THE CONDITIONAL EXTREME VALUE MODEL AND RELATED TOPICS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Bikramjit Das

August 2009

THE CONDITIONAL EXTREME VALUE MODEL AND RELATED TOPICS

Bikramjit Das , Ph.D.

Cornell University 2009

Extreme value theory (EVT) is often used to model environmental, financial and internet traffic data. Multivariate EVT assumes a multivariate domain of attraction condition for the distribution of a random vector necessitating that each component satisfy a marginal domain of attraction condition. Heffernan and Tawn [2004] and Heffernan and Resnick [2007] developed an approximation to the joint distribution of the random vector by conditioning on one of the components being in an extreme value domain. The usual method of analysis using multivariate extreme value theory often is not helpful either because of asymptotic independence or due to one component of the observation vector not being in a domain of attraction. These defects can be addressed by using the conditional extreme value model.

This thesis primarily concentrates on various aspects of this conditional extreme value model. Prior work left unresolved the consistency of different models obtained by conditioning on different components being extreme and we provide understanding of this issue. We also clarify the relationship between the conditional distributions and multivariate extreme value theory and extensions from one to the other. An important model issue is whether the limit measure in the conditional model is a product as this leads to a dichotomy in estimation of the model parameters. We propose three statistics which act as tools to detect the plausibility of using this model as well as whether the limit measure is a product or not.

This thesis also considers a graphical tool which has been in use for quite some time. The QQ plot is a commonly used technique for informally deciding whether a univariate random sample of size n comes from a specified distribution F. The QQ plot graphs the sample quantiles against the theoretical quantiles of F and then a visual check is made to see whether or not the points are close to a straight line. We consider the set $\boldsymbol{S}_n$ of points forming the QQ plot as a random closed set in $\mathbb{R}^2$. We show that under mild regularity conditions on the distribution F, $\boldsymbol{S}_n$ converges in probability to a closed, non-random set. In the heavy tailed case where $1-F$ is a regularly varying function, a similar result can be shown but a modification is necessary to provide a statistically sensible result since typically F is not completely known. This final technique is also used to marginally detect heavy-tails in the conditional extreme value model

BIOGRAPHICAL SKETCH

Bikramjit Das was born on April 1st, 1980 in Siliguri, India. He started preparatory school in the hills of Kalimpong and finished his secondary schooling in Ramakrishna Mission Vidyapith, Purulia. After completing his higher secondary education in St. Xavier's College, Kolkata, he joined the Statistics program in the Indian Statistical Institute, Kolkata where he received his Bachelors and Masters degree in Statistics in the years 2002 and 2004 respectively.

He joined the School of Operations Research and Information Engineering in Cornell University at Ithaca in the year 2004. Upon completion of his Ph. D. in August, 2009 he would be joining the Risklab, Department of Mathematics at ETH, Zürich as a post-doctoral researcher.

To my parents

ACKNOWLEDGEMENTS

First I would like to thank my advisor Professor Sidney Resnick for being an exceptional research guide. I have learnt many things from him through attending his courses as well as through personal discourses. He has been a pillar of encouragement and support throughout my graduate life at Cornell. I am also grateful for the generous financial support he provided through NSF Training grant for Graduate and Postdoctoral Training in Probability and its Applications, Cornell University and ARO contract W911NF-07-1-0078.

I would also like to thank my special committee comprising of Professors Bruce Turnbull and Philip Protter for their constant support and encouragement. I learnt a lot attending their classes and as their teaching assistants. A special thanks to Prof. Shane Henderson for acting as a proxy special committee member and providing helpful suggestions.

The faculty of the School of Operations Research and Information Engineering has been a constant source of encouragement. I would like to mention Professors David Shmyos, N.U. Prabhu, Shane Henderson for the nice courses they had offered which I had the opportunity to attend. I would especially mention Professors Gennady Samorodnitsky and Jan Rosinski whose special topics courses on infinitely divisible processes and long range dependence I enjoyed the most. I would like to thank Professor Francesca Mollinari for whom I came to learn many intricacies in the theory of random sets.

Prior to coming to Cornell, I had some wonderful professors from my undergraduate days at the Indian Statistical Institute without whose guidance and support I wouldn't have thought of pursuing a doctoral degree. Additionally I would like to thank my high school maths teacher S.L. Perumal who taught me the basics of Mathematics and Statistics.

On the social side Ithaca has been the nicest place to be a graduate student. I would like to thank all my friends who have made my years in Ithaca worth remembering. I would especially thank Ananda and Parthanil, who were my apartment mates for more than two years and had to bear with my eccentricities for that long; Ananda is the most cool tempered guy I have seen and a great friend to have, and Parthanil, my academic big brother had been ever-encouraging and helpful. And of course Amrita was an everlasting source of support through these final years in my graduate student life.

Finally, I would like to thank my family: Ma and Baba, who always believed in whatever I did, and without whom nothing close would be possible; Dadamoni, my big brother for having my back, always, in whatever I did; Boudi, Tatan and Titli for their endless love and best wishes through the years.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

The theory of extremes has often been used to understand the tail properties of multivariate probability distributions. Initial curiosity in extreme value theory can be traced backed to the use of measurements of maximum level of water-bodies in order to build dams. Subsequently, applications of this theory has been extensive in the areas of telecommunication, finance and environmental data. As an illustration, consider the various chemicals, particulate matter or biological material causing air pollution. Environmental agencies have specified numerical limits on the amount of these atmospheric pollutants (e.g., ozone, sulphur dioxide, carbon monoxide, etc.) being present in the air. Beyond these thresholds the said pollutants act as health hazards. This leads to the interest in the right tail behavior of the distribution of these pollutants, both marginally and jointly. Thus multivariate extreme value theory is relevant here. A nice exposition into the world of extreme value theory and heavy-tail analysis can be found in de Haan and Ferreira [2006], Resnick [2007, 2008b].

In the univariate set-up of extreme value theory, the distribution of a random variable is studied by assuming it to be in the maximal domain of attraction of some univariate extreme value distribution. Extending this to the multivariate case, which is of interest to us for this thesis, the extremal dependence structure between components is studied by assuming a multivariate maximal domain of attraction condition which requires that each marginal distribution belong to the maximal domain of attraction of some univariate extreme value distribution. The theory relies on centering and scaling the components appropriately and observing the limiting behavior near the tails of the distribution. A variety of concepts have been developed in order to understand this extremal dependence structure. Multivariate extreme value theory (abbreviated as MEVT, henceforth) provides a rich theory for extremal dependence in the case of asymptotic dependence [de Haan and Resnick, 1977, Resnick, 2008b, Pickands, 1981, de Haan and Ferreira, 2006] but fails to distinguish between asymptotic independence and actual independence. The extremal dependence structure in the asymptotically dependent case has been well-studied by Coles and Tawn [1991], de Haan and de Ronde [1998]. The idea of *coefficient of tail dependence* developed by Ledford and Tawn [1996, 1997, 1998] provided a better understanding of asymptotically independent behavior of various components and this concept has been elaborated with the help of *hidden regular variation*. See Resnick [2002], Maulik and Resnick [2005], Heffernan and Resnick [2005], Resnick [2008a] and [Resnick, 2007, Chapter 8].

Though theoretically elegant, in practice all the components of a random vector need not necessarily be in the domain of attraction of an extreme value distribution as required by the assumptions of MEVT. The natural way to proceed here would be assuming a subset of the random vector under consideration to be in a multivariate domain of attraction. The resulting model, called the conditional extreme value model, which is the focus of this thesis, can be related to MEVT and its related concepts like of asymptotic independence , hidden regular variation, etc.

1.1 Conditioned limit theory of extremes

This approach was studied in Heffernan and Tawn [2004] where the authors examined multivariate distributions by conditioning on one of the components being extreme. Their approach allowed a variety of examples of different types of asymptotic dependence and asymptotic independence in the sense of extreme value theory. Their statistical ideas were given a more mathematical framework by Heffernan and Resnick [2007] after some slight changes in assumptions to make the theory more probabilistically viable.

In their work, Heffernan and Resnick [2007] considered a bivariate random vector (X, Y) where the distribution of Y is in the domain of attraction of an extreme value distribution G_γ, where for $\gamma \in \mathbb{R}$,

$$G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}, \qquad 1 + \gamma x > 0. \tag{1.1.1}$$

For $\gamma = 0$, the distribution function is interpreted as $G_0(x) = e^{-e^{-y}}$, $y \in \mathbb{R}$. Instead of conditioning on Y being large, their theory was developed under the equivalent assumption of the existence of a vague limit for the modified joint distribution of the a suitably scaled and centered (X, Y). The vague convergence is in the space of Radon measures on $[-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)}$ where $\overline{\mathbb{E}}^{(\gamma)}, \gamma \in \mathbb{R}$ is the right closure of the set $\{x \in \mathbb{R} : 1 + \gamma x > 0\}$. The precise description of this vague limit is in Definition 2.1.1 in Chapter 2. It should be noted that this differs from the classical multivariate extreme value theory in the sense that only one of the marginal distributions is assumed to be in the domain of attraction of some univariate extreme value distribution. We call this model the *conditional extreme value model* and abbreviate it as CEV model for convenience.

Observe now that this model is not symmetric in the variables. In practice one may have a choice of variable to condition on being large and hence potentially different models are therefore possible. This raises the following issues: What is the relationship between these different models? Which of these should we choose? Since MEVT already exists as a coherent theory, it is of interest to understand the relationship of the CEV model to MEVT.

MEVT is often studied by first marginally transforming or standardizing into a measure which is *regularly varying on cones* of the Euclidean space. After the transformation, the *limit measure* provides all the information about the limit distribution of the random components. A similar concept of standardization for the CEV model is possible under some conditions. Such issues involving CEV modeling has been addressed in this thesis.

Any mathematical model would fail to gain relevance without empirical evidence. Can we observe a CEV model on real data? Heffernan and Tawn [2004] fitted their version of the conditional model to air pollution data. Internet traffic data is also potentially suited for CEV modeling. Information packets (data) are constantly being exchanged via the internet. Considering data being routed through a server, we can always observe the amount of data (in bytes), say S, being transferred between a pair of source and destination IP addresses in a specific session. Let the observed session duration be D and $R = S/D$ be the average rate of transfer. Prior empirical evidence strongly supports the belief that S is a heavy-tailed random variable, which means it is in an extreme value domain of attraction. Hence it is of interest to know whether (D, S) or (R, S) can be modeled as a CEV model or should we think in terms of classical MEVT. It would be useful to have tools to answer this question and this was the reason we considered detection statistics for the CEV model.

1.2 QQ plots: a graphical tool to detect heavy-tails

The QQ (or quantile-quantile) plot is a commonly used device to graphically, quickly and informally test the goodness-of-fit of a sample $X_1, \ldots, X_n$ to some distribution F in an exploratory way. The QQ plot measures how close the sample quantiles are to the theoretical quantile. Rather than considering individual quantiles, the QQ plot considers the sample as a whole and plots the sample quantiles against the theoretical quantiles of the specified target distribution F. It is intuitive and widely believed that the QQ plot should converge to a straight line as the sample size increases. Considering the QQ plot as a random closed set in $\mathbb{R}^2$, under mild regularity conditions on F, we show that the random set $\boldsymbol{S}_n$ converges in probability to a straight line (or some closed subset of a straight line), in a suitable topology on closed subsets of $\mathbb{R}^2$. We also show the asymptotic consistency of the slope of the least squares line through the QQ plot to the slope of the target straight line when the distribution F has bounded support and eventually extend these ideas to the case of heavy-tailed distributions.

In case of detecting the CEV model, one step is to marginally detect a distribution in an extreme value domain. The QQ plots provides a neat way to do this for the heavy-tailed case. Note that for heavy-tailed distribution, a sub-class of distributions in extreme value domains of attraction, the candidate distribution is unknown and hence a modified QQ plot is used.

1.3 Outline of Dissertation

In Chapters 2 and 3 we discuss the conditional extreme value model. Chapter 2 introduces the basics of the model and then we delve into various theoretical aspects of the model. We talk about the model consistency among the potentially different models available. We relate the CEV model to MEVT and connect the CEV model to regular variation on a cone in $\mathbb{R}^d$. We conclude with some illuminating examples of the model.

In Chapter 3 we look into more statistical aspects of the CEV model. We devise three statistics, namely Hillish, Pickandsish and Kendall's tau in order to detect the CEV model in a bivariate data set. In case the CEV model is validated, these statistics also detect whether the limit measure for the CEV model is in a product form or not which is important for modeling and estimation.

In Chapter 4 we prove convergence of QQ plots as random closed sets to a straight line. We then extend the result to heavy-tailed random variables where the target distribution is regularly varying but not known specifically. This technique is also used in marginally detecting a heavy-tailed random variable for the CEV model in Chapter 3.

CHAPTER 2

CONDITIONAL EXTREME VALUE MODEL: MODEL CONSISTENCY AND REGULAR VARIATION ON CONES

2.1 Introduction

In order to model air pollution data, Heffernan and Tawn [2004] came up with an approach to study multivariate distributions by conditioning on one of the components being in an extreme value domain of attraction. Their approach allowed a variety of examples of different types of asymptotic dependence and asymptotic independence in the sense of extreme value theory. Heffernan and Resnick [2007] after slightly modifying some of their assumptions made the model more probabilistically viable. They considered a bivariate random vector (X, Y) where the distribution of Y is in the domain of attraction of an extreme value distribution G_γ, where for $\gamma \in \mathbb{R}$,

$$G_\gamma(x) = \exp\{-(1+\gamma x)^{-1/\gamma}\}, \qquad 1 + \gamma x > 0. \tag{2.1.1}$$

For $\gamma = 0$, the distribution function is interpreted as $G_0(x) = e^{-e^{-y}}$, $y \in \mathbb{R}$. Instead of conditioning on Y being large, their theory was developed under the equivalent assumption of the existence of a vague limit for the modified joint distribution of the a suitably scaled and centered (X, Y). The vague convergence is in the space of Radon measures on $[-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)}$ where $\overline{\mathbb{E}}^{(\gamma)}, \gamma \in \mathbb{R}$ is the right closure of the set $\{x \in \mathbb{R} : 1 + \gamma x > 0\}$. The precise description of this vague limit is in Definition 2.1.1 below. It should be noted that this differs from the classical MEVT in the sense that only one of the marginal distributions is assumed to be in the domain of attraction of some univariate extreme value distribution.

In Section 2.2 of this chapter we study the consistency issues discussed in Heffernan and Tawn [2004] for such conditional models. In practice one may have a choice of variable to condition on being large and potentially different models are therefore possible. What is the relationship between these models? We show that if conditional approximations are possible no matter which variable as the conditioning variable, then in fact the joint distribution is in a classical multivariate domain of attraction of an extreme value law. A standard case is dealt with first and later extended to a general formulation. The relationship between multivariate extreme value theory and conditioned limit theory is discussed in Section 2.3. Conditions under which a standardized (to be defined appropriately) regular variation model can be used in place of the conditional model are discussed in this section. We also consider conditions under which the conditional extreme value model can be extended to the multivariate extreme value model in Section 2.4. Section 2.5 presents some illuminating examples to show the features of the conditional extreme value model.

2.1.1 Model setup and basic assumptions

The basic model set up for our discussion follows in the same lines as Heffernan and Resnick [2007]:

Definition 2.1.1 (Conditional extreme value model)**.** *Suppose that for random vector $(X, Y) \in \mathbb{R}^2$, we have $Y \sim F$. We make the following assumptions:*

1. *F is in the domain of attraction of an extreme value distribution, G_γ for some $\gamma \in \mathbb{R}$, as defined in* (1.1.1)*; that is, there exist functions $a(t) > 0, b(t) \in \mathbb{R}$ such*

that, as $t \to \infty$*, for* $1 + \gamma y > 0$*,*

$$t(1 - F(a(t)y + b(t))) = t\mathbf{P}\left(\frac{Y - b(t)}{a(t)} > y\right) \to (1 + \gamma y)^{-1/\gamma}. \tag{2.1.2}$$

2. *There exist functions* $\alpha(t) > 0$ *and* $\beta(t) \in \mathbb{R}$ *and a non-null Radon measure* μ *on Borel subsets of* $[-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)}$ *such that for each* $y \in \mathbb{E}^{(\gamma)}$*,*

$$[a] \quad t\mathbf{P}\left(\frac{X - \beta(t)}{\alpha(t)} \le x, \frac{Y - b(t)}{a(t)} > y\right) \to \mu([-\infty, x] \times (y, \infty]),$$

as $t \to \infty$ *for* (x, y) *continuity points of the limit.* (2.1.3)

$[b] \quad \mu([-\infty, x] \times (y, \infty])$ *is not a degenerate distribution in* x*,* (2.1.4)

$[c] \quad \mu([-\infty, x] \times (y, \infty]) < \infty.$ (2.1.5)

$[d] \quad H(x) := \mu([-\infty, x] \times (0, \infty])$ *is a probability distribution.* (2.1.6)

We say that (X, Y) *follows a* conditional extreme value model *(abbreviated as CEV model) if conditions (1) and (2) above are satisfied. We write* $(X, Y) \in CEV(\alpha, \beta, a, b, \gamma)$*.*

A non-null Radon measure $\mu(\cdot)$ is said to satisfy the *conditional non-degeneracy conditions* if both of (2.1.4) and (2.1.5) hold. Conditions (2.1.3),(2.1.4) and (2.1.5) imply that for continuity points (x, y) of $\mu(\cdot)$, as $t \to \infty$,

$$\mathbf{P}\left(\frac{X - \beta(t)}{\alpha(t)} \le x \,\middle|\, Y > b(t)\right) \to H(x) = \mu([-\infty, x] \times (0, \infty]), \tag{2.1.7}$$

i.e., a conditioned limit holds. Hence the name conditional extreme value model. Also note that (2.1.3) can be viewed in terms of vague convergence of measures in $\mathbb{M}_+\big([-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)}\big)$.

In this model the transformation $Y \mapsto Y^* = b^{\leftarrow}(Y)$ standardizes the Y-variable, i.e., we can assume $a^*(t) = t, b^*(t) = 0$. Hence a reformulation of

(2.1.3) leads to

$$tP\left(\frac{X-\beta(t)}{\alpha(t)} \le x, \frac{Y^*}{t} > y\right) \to \mu^*([-\infty,x]\times(y,\infty]), \quad \text{as } t\to\infty, \tag{2.1.8}$$

for (x,y) continuity points of μ^* where,

$$\mu^*([-\infty,x]\times(y,\infty]) = \begin{cases} \mu\big([-\infty,x]\times(\frac{y^\gamma-1}{\gamma},\infty]\big), & \text{if } \gamma\neq 0,\\ \mu\big([-\infty,x]\times(\log y,\infty]\big), & \text{if } \gamma = 0. \end{cases} \tag{2.1.9}$$

Under the above assumptions Heffernan and Resnick [2007] use a convergence to types argument to get properties of the normalizing and centering functions: there exists functions $\psi_1(\cdot), \psi_2(\cdot)$ such that

$$\lim_{t\to\infty}\frac{\alpha(tc)}{\alpha(t)} = \psi_1(c), \qquad \lim_{t\to\infty}\frac{\beta(tc)-\beta(t)}{\alpha(t)} = \psi_2(c). \tag{2.1.10}$$

This implies that $\psi(c) = c^\rho$ for some $\rho\in\mathbb{R}$ [de Haan and Ferreira, 2006, Theorem B.1.3]. ψ_2 can be either 0 or $\psi_2(c) = k\frac{c^\rho-1}{\rho}$ for some $c\neq 0$ [de Haan and Ferreira, 2006, Theorem B.2.1]. We refer often to these properties in Chapters 2 and 3 of this thesis.

2.1.2 Notation

We list below commonly used notation. The Appendix in Section A.1 contains information on regularly varying functions and extensions to such things as Π-varying functions, as well as a rapid review of vague convergence.

$\mathbb{R}_+^d$ $\quad [0,\infty)^d$.

$\overline{\mathbb{R}}_+^d$ $\quad [0,\infty]^d$. Also denote similarly $\overline{\mathbb{R}}^d = [-\infty,\infty]^d$.

$\mathbb{E}^*$	A nice subset of the compactified finite dimensional Euclidean space. Often denoted $\mathbb{E}$ with different subscripts and superscripts as required.
$\mathcal{E}^*$	The Borel $\sigma-$field of the subspace $\mathbb{E}^*$.
$\mathbb{M}_+(\mathbb{E}^*)$	The class of Radon measures on Borel subsets of $\mathbb{E}^*$.
$f^{\leftarrow}$	The left continuous inverse of a monotone function f. For an increasing function $f^{\leftarrow}(x) = \inf\{y : f(y) \geq x\}$. For a decreasing function $f^{\leftarrow}(x) = \inf\{y : f(y) \leq x\}$.
RV_ρ	The class of regularly varying functions with index ρ defined in (A.1.1).
Π	The function class Π reviewed in Section A.1.1 along with subclasses $\Pi_+(a(\cdot))$ and $\Pi_-\big(a(\cdot)\big)$ and *auxiliary function* $a(\cdot)$.
$\xrightarrow{v}$	Vague convergence of measures; see Section A.1.3.
G_γ	An extreme value distribution given by (1.1.1), with parameter $\gamma \in \mathbb{R}$, in the Von Mises parameterization.
$\mathbb{E}^{(\gamma)}$	$\{x : 1 + \gamma x > 0\}$ for $\gamma \in \mathbb{R}$.
$\overline{\mathbb{E}}^{(\gamma)}$	The closure on the right of the interval $\mathbb{E}^{(\gamma)}$.
$\overline{\overline{\mathbb{E}}}^{(\gamma)}$	The closure on both sides of the interval $\mathbb{E}^{(\gamma)}$.
$\mathbb{E}^{(\lambda,\gamma)}$	$\overline{\overline{\mathbb{E}}}^{(\lambda)} \times \overline{\overline{\mathbb{E}}}^{(\gamma)} \setminus \{(-\frac{1}{\lambda}, -\frac{1}{\gamma})\}$.
$D(G_\gamma)$	The domain of attraction of the extreme value distribution G_γ; i.e., the set of F's satisfying (2.1.2).For$\gamma > 0$, $F \in D(G_\gamma)$ is equivalent to $1 - F \in RV_{1/\gamma}$.

2.2 Consistency of CEV models

Suppose $(X, Y) \in \mathbb{R}^2$ satisfy conditions (2.1.3)-(2.1.6). Hence $(X, Y) \in CEV(\alpha, \beta, a, b, \gamma)$ with $F \in D(G_\gamma)$ where $Y \sim F$. Also assume (Y, X) now satisfy conditions (2.1.3)-(2.1.6), i.e., $(Y, X) \in CEV(c, d, \chi, \phi, \lambda)$ for some $\chi(t) > 0, c(t) > 0, \phi(t), d(t) \in \mathbb{R}, \lambda \in \mathbb{R}$ with $H \in D(G_\lambda)$ where $X \sim H$. What is the implication of these assumptions on the joint distribution of (X, Y) in the context of multivariate extreme value theory. We show in this section that these two assumptions indeed imply that (X, Y) is in the domain of attraction of a multivariate extreme value distribution.

2.2.1 Consistency: the standard case

Let us start with a simple case. Assume for (X, Y) as mentioned in the previous paragraph, the centering functions are all $zero$ and the norming functions are *identity functions*. Now set

$$\mathbb{E} \ = [0, \infty]^2 \setminus \{\mathbf{0}\}, \qquad \mathbb{E}_0 = (0, \infty] \times (0, \infty],$$
$$\mathbb{E}_{\sqcap} = [0, \infty] \times (0, \infty], \qquad \mathbb{E}_{\sqsupset} = (0, \infty] \times [0, \infty].$$

Figure 2.1 illustrates these four types of cones in two dimensions.

Before proceeding, we review the definition of multivariate regular variation on cones ([Resnick, 2007, page 173], Resnick [2008a], Davydov et al. [2007]).

Definition 2.2.1. *$\mathfrak{C} \subset \overline{\mathbb{R}}^d$ is a cone if $x \in \mathfrak{C}$ implies that $tx \in \mathfrak{C}$ for any $t > 0$. Now a $d-$dimensional random vector $\boldsymbol{Z} \in \mathbb{R}^d$ is* multivariate regularly varying *on cone $\mathfrak{C}$ in $\overline{\mathbb{R}}^d$ if there exists a function $b(t) \to \infty$ and a non-null Radon measure ν on $\mathfrak{C}$ such*

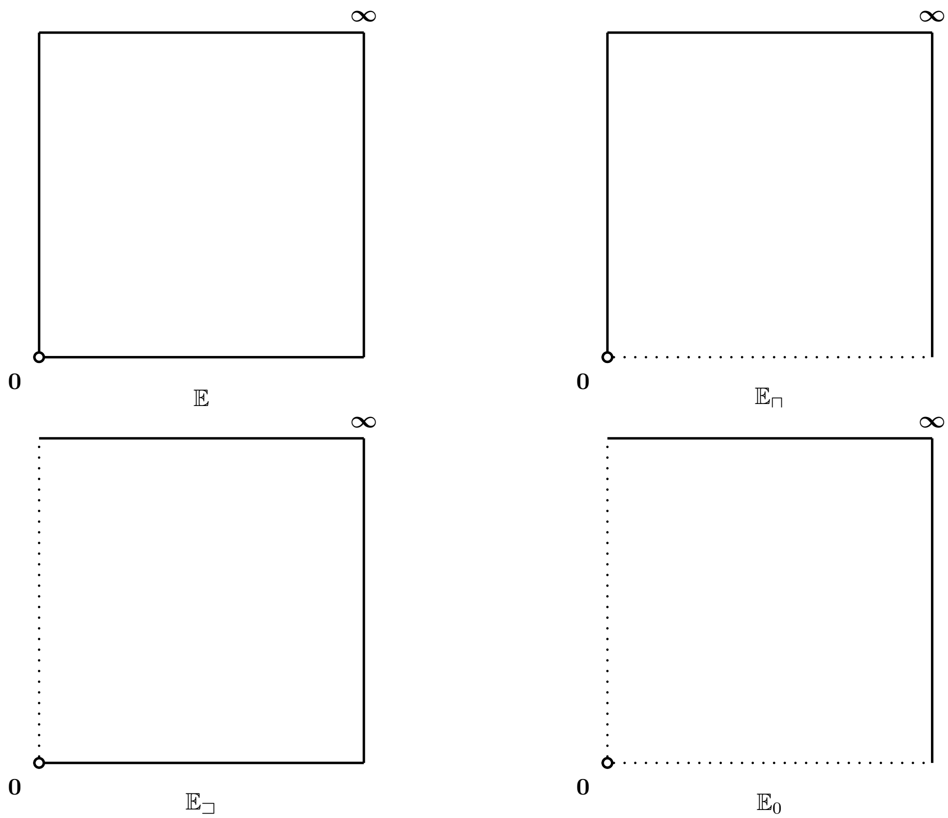

Figure 2.1: Different cones in 2-dimensions

that, as $t \to \infty$,

$$t\mathbf{P}\left[\frac{\boldsymbol{Z}}{b(t)} \in \cdot\right] \xrightarrow{v} \nu(\cdot), \text{ in } \mathbb{M}_+(\mathfrak{C}). \tag{2.2.1}$$

Remark 2.2.1. It follows from (2.2.1) that the limit measure $\nu(\cdot)$ satisfies the homogeneity property that for a relatively compact set $B \subset \mathfrak{C}$,

$$\nu(cB) = c^{-\alpha}\nu(B) \qquad c > 0, \tag{2.2.2}$$

for some $\alpha > 0$. This also implies that $b(\cdot)$ is regularly varying with index $1/\alpha$.

Remark 2.2.2. The regular variation in Definition 2.2.1 is *standard* if $\boldsymbol{Z} \in \mathbb{R}_+^d$ and $b(t) \equiv t$. In this case we have equation (2.2.2) with $\alpha = 1$.

Note that $\mathbb{E}, \mathbb{E}_{\sqcap}, \mathbb{E}_{\sqsupset}, \mathbb{E}_0$ are all cones in $\overline{\mathbb{R}}^2$. The following result shows that standard regular variation on both $\mathbb{E}_{\sqcap}$ and $\mathbb{E}_{\sqsupset}$ imply standard regular variation on the bigger cone $\mathbb{E}_{\sqcap} \cup \mathbb{E}_{\sqsupset} = \mathbb{E}$. This is the introduction to the more general consistency results in the CEV model.

Theorem 2.2.1. *Suppose we have a bivariate random vector $(X, Y) \in \mathbb{R}_+^2$. Now assume that*

$$t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}_{\sqcap}\big), \tag{2.2.3}$$

$$t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \nu(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}_{\sqsupset}\big). \tag{2.2.4}$$

where both μ and ν satisfy appropriate conditional non-degeneracy conditions corresponding to (2.1.4)–(2.1.6). *Then (X, Y) is standard regularly varying in $\mathbb{E}$, i.e.,*

$$t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} (\mu \diamond \nu)(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}\big) \tag{2.2.5}$$

where $(\mu \diamond \nu)$ is a Radon measure on $\mathbb{E}$ such that

$$(\mu \diamond \nu)|_{\mathbb{E}_{\sqcap}}(\cdot) = \mu(\cdot) \quad \text{on } \mathbb{E}_{\sqcap} \qquad \text{and} \qquad (\mu \diamond \nu)|_{\mathbb{E}_{\sqsupset}}(\cdot) = \nu(\cdot) \quad \text{on } \mathbb{E}_{\sqsupset}.$$

Proof. First, note that if A is relatively compact in $\mathbb{E}_0 = \mathbb{E}_{\sqcap} \cap \mathbb{E}_{\sqsupset}$ with $\mu(\partial A) = 0 = \nu(\partial A)$, we will have

$$\mu(A) = \lim_{t \to \infty} t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in A\Big] = \nu(A) \tag{2.2.6}$$

(Portmanteau Theorem for vague convergence, Theorem (3.2) in Resnick [2007]). Hence we have a unique vague limit on $M_+(\mathbb{E}_0)$; i.e.,

$$\mu(\cdot)|_{\mathbb{E}_{\sqcap} \cap \mathbb{E}_{\sqsupset}} = \nu(\cdot)|_{\mathbb{E}_{\sqcap} \cap \mathbb{E}_{\sqsupset}}. \tag{2.2.7}$$

For $\epsilon > 0$ let us define:

$$B_1^\epsilon = [0, \epsilon) \times [\epsilon, \infty] \in \mathcal{E}_{\sqcap},$$

$$B_2^\epsilon = [\epsilon, \infty] \times [0, \infty] \in \mathcal{E}_{\sqsupset}.$$

We will prove the theorem with the following two claims.

Claim 2.2.1. *Let $A \in \mathcal{E}$ be a relatively compact set in $\mathbb{E}$. Define the set*

$$[(\mu \diamond \nu)(A)] = \{\mu(A \cap B_1^\epsilon) + \nu(A \cap B_2^\epsilon) : 0 < \epsilon < d(0, A)/\sqrt{2}\}.$$

Then $[(\mu \diamond \nu)(A)]$ is a singleton set and if we denote the unique element of this set by $(\mu \diamond \nu)(A)$, then $(\mu \diamond \nu)(\cdot)$ is a Radon measure on $\mathbb{E}$.

Proof. Since A is relatively compact, $d(0, A) > 0$. Take $0 < \epsilon < \delta < d(0, A)/\sqrt{2}$. Now to show that $[(\mu \diamond \nu)(A)]$ is a singleton, it suffices to check the following:

$$\mu(A \cap B_1^\epsilon) + \nu(A \cap B_2^\epsilon) = \mu(A \cap B_1^\delta) + \nu(A \cap B_2^\delta).$$

Let us define

$$F_1 = B_1^\epsilon \cap B_1^\delta, \qquad F_2 = B_1^\epsilon \cap (B_1^\delta)^c,$$
$$F_3 = (B_1^\epsilon)^c \cap B_1^\delta, \qquad F_4 = B_2^\epsilon \cap (B_1^\delta)^c \cap (B_2^\delta)^c,$$
$$F_5 = B_2^\delta.$$

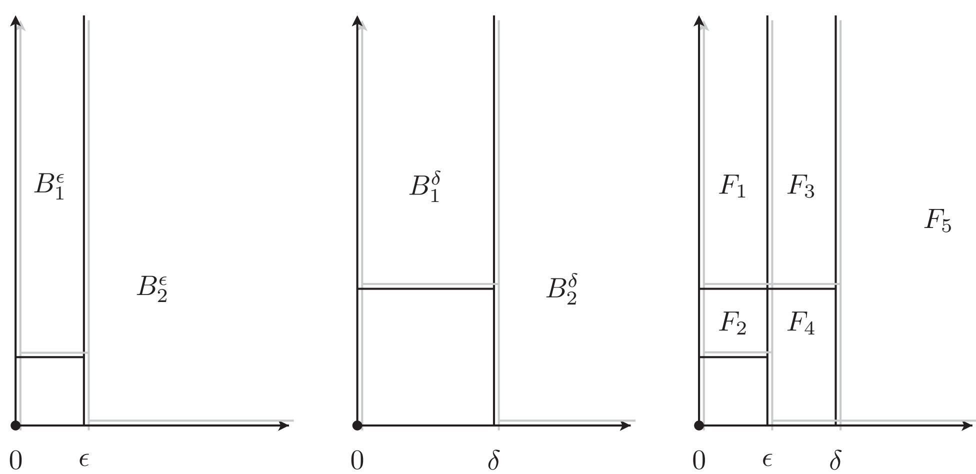

Figure 2.2: Different partitions of $\mathbb{E}_0$

Referring to Figure 2.2 we have

$$B_1^\epsilon = F_1 \cup F_2, \qquad B_1^\delta = F_1 \cup F_3,$$
$$B_2^\epsilon = F_3 \cup F_4 \cup F_5 \qquad B_2^\delta = F_5.$$

First note that $F_1, F_2, F_3 \in \mathcal{E}_\sqcap$, $F_3, F_4, F_5 \in \mathcal{E}_\sqsupset$. Hence $A \cap F_3 \in \mathcal{E}_0$ and from (2.2.7) we have

$$\mu(A \cap F_3) = \nu(A \cap F_3).$$

Also note that $\delta < d(0, A)/\sqrt{2}$ implies $A \subset B_1^\delta \cup B_2^\delta = F_1 \cup F_3 \cup F_5$. So

$$A \cap F_2 = \emptyset, A \cap F_4 = \emptyset. \tag{2.2.8}$$

Now

$$\begin{aligned}
&\mu(A \cap B_1^\epsilon) + \nu(A \cap B_2^\epsilon) \\
&= \mu(A \cap (F_1 \cup F_2)) + \nu(A \cap (F_3 \cup F_4 \cup F_5)) \\
&= \mu(A \cap F_1) + \mu(A \cap F_2) + \nu(A \cap F_3) + \nu(A \cap F_4) + \nu(A \cap F_5) \\
&= \mu(A \cap F_1) + \mu(A \cap F_3) + \nu(A \cap F_5) \qquad \text{(using (2.2.7) and (2.2.8))} \\
&= \mu(A \cap (F_1 \cup F_3)) + \nu(A \cap F_5) \\
&= \mu(A \cap B_1^\delta) + \nu(A \cap B_2^\delta).
\end{aligned}$$

Thus $[(\mu \diamond \nu)(A)]$ is a singleton and we denote its unique element by $(\mu \diamond \nu)(A)$. It is easy to see that $(\mu \diamond \nu)(\cdot)$ is a Radon measure. $\square$

Claim 2.2.2. *Let A be a relatively compact set in $\mathbb{E}$ with $(\mu \diamond \nu)(\partial A) = 0$. Then*

$$t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in A\Big] \to (\mu \diamond \nu)(A).$$

Proof. Choose $0 < \epsilon < d(0, A)$ s.t. $\mu(\partial B_1^\epsilon) = \nu(\partial B_2^\epsilon) = 0$. Now, $\partial(A \cap B_1^\epsilon) \subseteq \partial A \cup \partial B_1^\epsilon$. Therefore

$$\begin{aligned}
(\mu \diamond \nu)(\partial(A \cap B_1^\epsilon)) &\leq (\mu \diamond \nu)(\partial A) + (\mu \diamond \nu)(\partial B_1^\epsilon) \\
&= 0 + \mu(\partial B_1^\epsilon) = 0 \qquad \text{(using Claim 2.2.1 since } \partial B_1^\epsilon \in \mathcal{E}_\sqcap).
\end{aligned}$$

Thus $A \cap B_1^\epsilon$ is relatively compact in $\mathbb{E}_\sqcap$ and $\mu(\partial(A \cap B_1^\epsilon)) = 0$. Similarly $A \cap B_2^\epsilon$ is relatively compact in $\mathbb{E}_\sqsupset$ and $\nu(\partial(A \cap B_2^\epsilon)) = 0$. Hence we have

$$\begin{aligned} t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in A\Big] &= t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in A \cap B_1^\epsilon\Big] + t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in A \cap B_2^\epsilon\Big] \\ &\to \mu(A \cap B_1^\epsilon) + \nu(A \cap B_2^\epsilon) \\ &= (\mu \diamond \nu)(A). \end{aligned}$$

Hence the claim. $\square$

This claim holds for any $A \in \mathcal{E}$ which is relatively compact with $\mu(\partial A) = 0$. Thus from Portmanteau Theorem for vague convergence [Resnick, 2007, Theorem (3.2)] we have proven Claim 2.2.2.

Now we show that $(\mu \diamond \nu)|_{\mathbb{E}_\sqcap}(\cdot) = \mu(\cdot)$ on $\mathbb{E}_\sqcap$. If $A \in \mathcal{E}_\sqcap$, choose any $0 < \epsilon < d(0, A)/\sqrt{2}$ for defining $(\mu \diamond \nu)(A)$. Now we have $A \cap B_2^\epsilon = A \cap B_2^\epsilon \cap \mathbb{E}_\sqcap \in \mathcal{E}_0$. Therefore

$$\begin{aligned} (\mu \diamond \nu)(A) &= \mu(A \cap B_1^\epsilon) + \nu(A \cap B_2^\epsilon) \\ &= \mu(A \cap B_1^\epsilon) + \nu(A \cap B_2^\epsilon \cap \mathbb{E}_\sqcap) \end{aligned}$$

and from (2.2.6), since $A \cap B_2^\epsilon \cap \mathbb{E}_\sqcap \in \mathcal{E}_0$, this is

$$\begin{aligned} &= \mu(A \cap B_1^\epsilon) + \mu(A \cap B_2^\epsilon \cap \mathbb{E}_\sqcap) \\ &= \mu(A \cap (B_1^\epsilon \cup (B_2^\epsilon \cap \mathbb{E}_\sqcap))) = \mu(A). \end{aligned}$$

We can prove $(\mu \diamond \nu)|_{\mathbb{E}_\sqsupset}(\cdot) = \nu(\cdot)$ on $\mathbb{E}_\sqsupset$ similarly. $\square$

A more general result is stated and proved next. Note, however, that the proof for Theorem 2.2.1 is relatively easy because of the standard case assumptions and is instructive to read.

2.2.2 Consistency: the general case

Thus we see that multivariate regular variation on both the cones $\mathbb{E}_{\sqsupset}$ and $\mathbb{E}_{\sqcap}$ implies multivariate regular variation on the larger cone $\mathbb{E}_{\sqsupset} \cup \mathbb{E}_{\sqcap} = \mathbb{E}$. Now we will discuss the general situation in which each marginal distribution is in the domain of attraction of an extreme value distribution. Recall our notation, $\mathbb{E}^{(\gamma)} = \{x \in \mathbb{R} : 1 + \gamma x > 0\}$ for $\gamma \in \mathbb{R}$. We denote $\overline{\mathbb{E}}^{(\gamma)}$ to be the right closure of $\mathbb{E}^{(\gamma)}$, i.e.,

$$\overline{\mathbb{E}}^{(\gamma)} = \begin{cases} (-\frac{1}{\gamma}, \infty] & \gamma > 0 \\ (-\infty, \infty] & \gamma = 0 \\ (-\infty, -\frac{1}{\gamma}] & \gamma < 0. \end{cases} \tag{2.2.9}$$

$\overline{\overline{\mathbb{E}}}^{(\gamma)}$ denotes the set we get by closing $\mathbb{E}^{(\gamma)}$ on both sides. Also denote $\mathbb{E}^{(\lambda,\gamma)} := \overline{\overline{\mathbb{E}}}^{(\lambda)} \times \overline{\overline{\mathbb{E}}}^{(\gamma)} \setminus \{(-\frac{1}{\lambda}, -\frac{1}{\gamma})\}$.

Theorem 2.2.2. *Suppose we have a bivariate random vector $(X, Y) \in \mathbb{R}^2$ and non-negative functions $\alpha(\cdot), a(\cdot)$, $\chi(\cdot), c(\cdot)$ and real valued functions $\beta(\cdot), b(\cdot), \phi(\cdot), d(\cdot)$ such that*

$$t\mathbf{P}\Big[\Big(\frac{X-\beta(t)}{\alpha(t)}, \frac{Y-b(t)}{a(t)}\Big) \in \cdot\,\Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+\big([-\infty,\infty] \times \overline{\mathbb{E}}^{(\gamma)}\big), \tag{2.2.10}$$

$$t\mathbf{P}\Big[\Big(\frac{X-\phi(t)}{\chi(t)}, \frac{Y-d(t)}{c(t)}\Big) \in \cdot\,\Big] \xrightarrow{v} \nu(\cdot) \text{ in } \mathbb{M}_+\big(\overline{\mathbb{E}}^{(\lambda)} \times [-\infty,\infty]\big) \tag{2.2.11}$$

for some $\lambda, \gamma \in \mathbb{R}$, where both μ and ν satisfy the appropriate conditional non-degeneracy conditions corresponding to (2.1.4) *and* (2.1.5). *Then (X, Y) is in the domain of attraction of a multivariate extreme value distribution on $\mathbb{E}^{(\lambda,\gamma)}$ in the following sense:*

$$t\mathbf{P}\Big[\Big(\frac{X-\phi(t)}{\chi(t)}, \frac{Y-b(t)}{a(t)}\Big) \in \cdot\,\Big] \xrightarrow{v} (\mu \diamond \nu)(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}^{(\lambda,\gamma)}\big)$$

where $(\mu \diamond \nu)(\cdot)$ is a non-null Radon measure on $\mathbb{E}^{(\lambda,\gamma)}$.

Proof. Let us assume that $\lambda > 0, \gamma > 0$ first. The other cases can be dealt with similarly. It should be noted now, that (2.2.10) and (2.2.11) respectively imply that

$$t\mathbf{P}\Big(\frac{Y-b(t)}{a(t)} > y\Big) \to (1+\gamma y)^{-1/\gamma}, \quad 1+\gamma y > 0, \tag{2.2.12}$$

$$t\mathbf{P}\Big(\frac{X-\phi(t)}{\chi(t)} > x\Big) \to (1+\lambda x)^{-1/\lambda}, \quad 1+\lambda x > 0. \tag{2.2.13}$$

Hence for $(x,y) \in \mathbb{E}^{(\lambda)} \times \mathbb{E}^{(\gamma)}$,which are continuity points of the limit measures μ and ν,

$$\begin{aligned} Q_t(x,y) :=& t\mathbf{P}\Big[\Big(\frac{X-\phi(t)}{\chi(t)}, \frac{Y-b(t)}{a(t)}\Big) \in \big([-\infty,x]\times[-\infty,y]\big)^c\Big] \\ =& t\mathbf{P}\Big[\frac{X-\phi(t)}{\chi(t)} > x\Big] + t\mathbf{P}\Big[\frac{Y-b(t)}{a(t)} > y\Big] \\ &- t\mathbf{P}\Big[\frac{X-\phi(t)}{\chi(t)} > x, \frac{Y-b(t)}{a(t)} > y\Big] \\ =& A_t(x) + B_t(y) - C_t(x,y) \text{ (say)}. \end{aligned} \tag{2.2.14}$$

If we can show that $Q_t(x,y)$ has a limit and the limit is non-degenerate over (x,y) then we are done. As $t\to\infty$ we have the limits for $A_t(x)$ and $B_t(y)$ from equations (2.2.13) and (2.2.12) respectively. Clearly $0 \le C_t(x,y) \le \min(A_t(x), B_t(y))$ and these inequalities would hold for any limit of Q_t as well.

From [Heffernan and Resnick, 2007, Proposition 1], there exist functions $\psi_1(\cdot), \psi_2(\cdot), \psi_3(\cdot), \psi_4(\cdot)$ such that for $z > 0$,

$$\lim_{t\to\infty}\frac{\alpha(tz)}{\alpha(t)} = \psi_1(z) = z^{\rho_1}, \qquad \lim_{t\to\infty}\frac{\beta(tz)-\beta(t)}{\alpha(t)} = \psi_2(z), \tag{2.2.15}$$

$$\lim_{t\to\infty}\frac{c(tz)}{c(t)} = \psi_3(z) = z^{\rho_2}, \qquad \lim_{t\to\infty}\frac{d(tz)-d(t)}{c(t)} = \psi_4(z). \tag{2.2.16}$$

for some real constants ρ_1 and ρ_2. Assume ρ_1 and ρ_2 to be positive for the time being. Here either $\psi_2(z) = 0$ which would imply $\lim\limits_{t\to\infty}\frac{\beta(t)}{\alpha(t)} = 0$ (from [Bingham et al., 1987, Theorem 3.1.12 a,c]) or we can have $\psi_2(z) = k\frac{z^{\rho_1}-1}{\rho_1}$ for some $k \neq 0$, which means $\lim_{t\to\infty} \frac{\beta(t)}{\alpha(t)} = \frac{k}{\rho_1}$ ([de Haan and Ferreira, 2006, Proposition B.2.2]). Hence allowing the constant k to be zero as well, we can write both cases as $\lim_{t\to\infty} \frac{\beta(t)}{\alpha(t)} = \frac{k_1}{\rho_1}$ for some $k_1 \in \mathbb{R}$. Similarly we have $\lim_{t\to\infty} \frac{d(t)}{c(t)} = \frac{k_2}{\rho_2}$ for some $k_2 \in \mathbb{R}$.

Additionally, from the marginal domain of attraction conditions for X, Y we have

$$\lim_{t\to\infty} \frac{b(tz) - b(t)}{a(t)} = \frac{z^\gamma - 1}{\gamma}, \qquad \text{for } y > 0,$$

$$\text{which implies} \qquad \lim_{t\to\infty} \frac{a(tz)}{a(t)} = z^\gamma, \tag{2.2.17}$$

and

$$\lim_{t\to\infty} \frac{\phi(tw) - \phi(t)}{\chi(t)} = \frac{w^\lambda - 1}{\lambda}, \qquad \text{for } w > 0,$$

$$\text{which implies} \qquad \lim_{t\to\infty} \frac{\chi(tw)}{\chi(t)} = w^\lambda. \tag{2.2.18}$$

Observe that

$$\begin{aligned} C_t(x,y) &= t\mathbf{P}\Big[\frac{X - \phi(t)}{\chi(t)} > x, \frac{Y - b(t)}{a(t)} > y\Big] \\ &= t\mathbf{P}\Big[\frac{X - \beta(t)}{\alpha(t)} > \Big(x + \frac{\phi(t)}{\chi(t)}\Big)\frac{\chi(t)}{\alpha(t)} - \frac{\beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)} > y\Big]. \end{aligned} \tag{2.2.19}$$

We can also write

$$C_t(x,y) = t\mathbf{P}\Big[\frac{X - \phi(t)}{\chi(t)} > x, \frac{Y - d(t)}{c(t)} > \Big(y + \frac{b(t)}{a(t)}\Big)\frac{a(t)}{c(t)} - \frac{d(t)}{c(t)}\Big]. \tag{2.2.20}$$

From [de Haan and Ferreira, 2006, Proposition B.2.2] we have that

$$\frac{b(t)}{a(t)} \to \frac{1}{\gamma} \quad \text{and} \quad \frac{\phi(t)}{\chi(t)} \to \frac{1}{\lambda}. \tag{2.2.21}$$

We analyze $C_t(x,y)$ for the different cases now. First we will show that at least one of the limits $\lim_{t\to\infty} \frac{\chi(t)}{\alpha(t)}$ and $\lim_{t\to\infty} \frac{a(t)}{c(t)}$ has to exist. Suppose both do not exist. We have for $(x, y) \in \mathbb{E}^{(\lambda)} \times \mathbb{E}^{(\gamma)}$, which are continuity points of the limit measures μ and ν,

$$t\mathbf{P}\Big[\frac{X-\beta(t)}{\alpha(t)} > x, \frac{Y-b(t)}{a(t)} > y\Big] \to \mu((x,\infty] \times (y,\infty]), \tag{2.2.22}$$

$$t\mathbf{P}\Big[\frac{X-\phi(t)}{\chi(t)} > x, \frac{Y-d(t)}{c(t)} > y\Big] \to \nu((x,\infty] \times (y,\infty]). \tag{2.2.23}$$

Now (2.2.22) implies that

$$t\mathbf{P}\Big[\frac{X-\phi(t)}{\chi(t)}\frac{\chi(t)}{\alpha(t)} + \frac{\phi(t)-\beta(t)}{\alpha(t)} > x, \frac{Y-d(t)}{c(t)}\frac{c(t)}{a(t)} + \frac{d(t)-b(t)}{a(t)} > y\Big]$$
$$\to \mu\big((x,\infty] \times (y,\infty]\big)$$

which is equivalent to

$$t\mathbf{P}\Big[\frac{X-\phi(t)}{\chi(t)} > \frac{\alpha(t)}{\chi(t)}\big(x - \frac{\phi(t)-\beta(t)}{\alpha(t)}\big), \frac{Y-d(t)}{c(t)} > \frac{a(t)}{c(t)}\big(y - \frac{d(t)-b(t)}{a(t)}\big)\Big]$$
$$\to \mu\big((x,\infty] \times (y,\infty]\big).$$

From (2.2.23) we also have that the left side of the previous line has a limit

$$t\mathbf{P}\Big[\frac{X-\phi(t)}{\chi(t)} > \frac{\alpha(t)}{\chi(t)}\big(x - \frac{\phi(t)-\beta(t)}{\alpha(t)}\big), \frac{Y-d(t)}{c(t)} > \frac{a(t)}{c(t)}\big(y - \frac{d(t)-b(t)}{a(t)}\big)\Big]$$
$$\to \nu\big((f(x),\infty] \times (g(y),\infty]\big)$$

for some $(f(x), g(y))$, assumed to be a continuity point of the limit ν, iff as $t \to \infty$, the following two limits hold,

$$\frac{\alpha(t)}{\chi(t)}\Big(x - \frac{\phi(t)-\beta(t)}{\alpha(t)}\Big) \to f(x), \tag{2.2.24}$$

$$\frac{a(t)}{c(t)}\Big(y - \frac{d(t)-b(t)}{a(t)}\Big) \to g(y). \tag{2.2.25}$$

For ν to be non-degenerate f and g should be non-constant and we also have $\mu\big((x,\infty] \times (y,\infty]\big) = \nu\big((f(x),\infty] \times (g(y),\infty]\big)$. Considering (2.2.24) and (2.2.25) we can see that the limit as $t \to \infty$ exists if and only if $\lim\limits_{t\to\infty} \frac{a(t)}{c(t)}$ and $\lim\limits_{t\to\infty} \frac{\chi(t)}{\alpha(t)}$ exists.

We conclude $\lim\limits_{t\to\infty} \frac{\chi(t)}{\alpha(t)} \in [0,\infty]$ and consider the following cases.

- **Case 1:** $\lim_{t\to\infty} \frac{\chi(t)}{\alpha(t)} = \infty$.

  Consider (2.2.19) and note

$$\Big(x + \frac{\phi(t)}{\chi(t)}\Big)\frac{\chi(t)}{\alpha(t)} - \frac{\beta(t)}{\alpha(t)} \to \Big(x + \frac{1}{\lambda}\Big) \times \infty - \frac{k_1}{\rho_1} = \infty,$$

  which entails

$$\lim_{t\to\infty} C_t(x, y) = \mu(\{\infty\} \times (y, \infty]) = 0.$$

  Hence

$$\lim_{t\to\infty} Q_t(x, y) = (1 + \lambda x)^{-1/\lambda} + (1 + \gamma y)^{-1/\gamma}.$$

- **Case 2:** $\lim_{t\to\infty} \frac{\chi(t)}{\alpha(t)} = M \in (0, \infty)$.

  Again from (2.2.19), we have

$$\Big(x + \frac{\phi(t)}{\chi(t)}\Big)\frac{\chi(t)}{\alpha(t)} - \frac{\beta(t)}{\alpha(t)} \to \Big(x + \frac{1}{\lambda}\Big) \times M - \frac{k_1}{\rho_1} = f(x) \text{ (say)}.$$

  Therefore

$$\lim_{t\to\infty} C_t(x, y) = \mu((f(x), \infty] \times (y, \infty]) \leq (1 + \lambda y)^{-1/\lambda}$$

  with strict inequality holding for some x because of the non-degeneracy condition (2.1.4) for μ. Hence

$$\lim_{t\to\infty} Q_t(x, y) = ((1 + \lambda x)^{-1/\lambda} + (1 + \gamma y)^{-1/\gamma} - \mu((f(x), \infty] \times (y, \infty]).$$

- **Case 3:** $\lim_{t\to\infty} \frac{\chi(t)}{\alpha(t)} = 0$.

  In this case (2.2.19) leads to a degenerate limit in x for $C_t(x, y)$ and putting $M_1 = \frac{k}{\rho_1}$ we get

$$\lim_{t\to\infty} C_t(x, y) = \mu((M_1, \infty] \times (y, \infty]) =: f_1(y) \leq (1 + \gamma y)^{-1/\gamma}.$$

  So consider (2.2.20).

  1. If $\lim_{t\to\infty} \frac{a(t)}{c(t)}$ exists in $(0, \infty]$, then we can use a similar technique as in case 1 or case 2 to obtain a non-degenerate limit for $Q_t(x, y)$.

2. Assume $\lim_{t\to\infty} \frac{a(t)}{c(t)} = 0$. Then for some $M_2 \in \mathbb{R}$,

$$\lim_{t\to\infty} C_t(x,y) = \nu((x,\infty] \times (M_2,\infty] =: f_2(x) \leq (1+\lambda x)^{-1/\lambda}$$

Therefore we have for any $(x,y) \in \mathbb{E}^{(\lambda)} \times \mathbb{E}^{(\gamma)}$,which are continuity points of the limit measures μ and ν,

$$f_1(y) = \mu((M_1,\infty] \times (y,\infty]) = \nu((x,\infty] \times (M_2,\infty]) = f_2(x).$$

It is easy to check now that for any $(x,y) \in \mathbb{E}^{(\lambda)} \times \mathbb{E}^{(\gamma)}$, which are continuity points of the limit measures μ and ν, we have $f_1(y) = f_2(x) = 0$. Hence $C_t(x,y) \to 0$ and thus $Q_t(x,y)$ has a non-degenerate limit.

This proves the result. □

Remark 2.2.3. The significant feature of Theorem 2.2.2 is that we do not need any other condition on the normalizing functions. The convergences (2.2.10)and (2.2.11) imply that $\alpha = O(\chi)$ and $c = O(a)$ as $t \to \infty$. If either $\alpha = o(\chi)$ or $c = o(a)$ then we have asymptotic independence and the existence of hidden regular variation.

2.2.3 Consistency: the absolutely continuous case

It is instructive to consider the consistency issue when (X,Y) has a joint density since calculations become more explicit. We state the consistency result particularizing Theorem 2.2.2; the proof for the standard case is provided later.

Proposition 2.2.3. *Assume g_ρ denotes the density of an extreme value distribution G_ρ with shape parameter $\rho \in \mathbb{R}$ in* (1.1.1) *and that $(X,Y) \in \mathbb{R}^2$ is a bivariate random vector. We suppose*

1. (X, Y) *has a density* $f_{X,Y}(x, y)$.

2. *The marginal densities* $f_X,\ f_Y$ *satisfy (as* $t \to \infty$*):*

$$t\chi(t) f_X(c(t)x + d(t)) \to g_\lambda(x), x \in \mathbb{E}^{(\lambda)}, \tag{2.2.26}$$

$$ta(t) f_Y(a(t)y + b(t)) \to g_\gamma(y), y \in \mathbb{E}^{(\gamma)}. \tag{2.2.27}$$

3. *The joint density satisfies (as* $t \to \infty$*):*

$$\begin{aligned} &t\alpha(t)a(t) f_{X,Y}(\alpha(t)x + \beta(t), a(t)y + b(t)) \\ &\qquad\qquad\qquad \to g_1(x, y) \in L^1([-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)}), \end{aligned} \tag{2.2.28}$$

$$\begin{aligned} &t\chi(t)d(t) f_{X,Y}(\chi(t)x + \phi(t), c(t)y + d(t)) \\ &\qquad\qquad\qquad \to g_2(x, y) \in L^1(\overline{\mathbb{E}}^{(\lambda)} \times [-\infty, \infty]), \end{aligned} \tag{2.2.29}$$

where $g_1(x, y), g_2(x, y) \geq 0$ *are non-trivial,* 0 *outside of* $[-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)}$ *and* $\overline{\mathbb{E}}^{(\lambda)} \times [-\infty, \infty]$ *respectively.*

Then

$$t\mathbf{P}\Big[\Big(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}^{(\lambda,\gamma)}\big),$$

for some non-degenerate Radon measure μ *on* $\mathbb{E}^{(\lambda,\gamma)}$.

Example 2.2.1. The following provides an example of the previous two propositions (restricted to the non-negative orthant). Suppose (X, Y) is a bivariate random variable with joint density

$$f_{X,Y}(x, y) = \frac{4x}{(x^2 + y)^3} + \frac{4y}{(x + y^2)^3}, \quad x \geq 1, y \geq 1.$$

Clearly the conditions of Proposition 2.2.3 hold: As $t \to \infty$,

$$\begin{aligned} &t^2 f_X(tx) \to \frac{2}{x^2}, \quad t^2 f_Y(ty) \to \frac{2}{y^2}, \quad x, y > 0, \\ &t^{5/2} f_{X,Y}(tx, \sqrt{t}y) \to \frac{4y}{(x + y^2)^3} =: g_1(x, y) \in L_1(\mathbb{E}_{\sqsupset}) \\ &t^{5/2} f_{X,Y}(\sqrt{t}x, ty) \to \frac{4x}{(x^2 + y)^3} =: g_1(x, y) \in L_1(\mathbb{E}_{\sqcap}). \end{aligned}$$

Now it is easy to observe the following as $t \to \infty$:

$$
\begin{aligned}
&t\mathbf{P}(\frac{X}{\sqrt{t}} \le x, \frac{Y}{t} > y) \to \frac{1}{y} - \frac{1}{y+x^2}, \quad x \ge 0, y > 0,\\
&t\mathbf{P}(\frac{X}{t} > x, \frac{Y}{\sqrt{t}} \le y) \to \frac{1}{x} - \frac{1}{x+y^2}, \quad x > 0, y \ge 0,\\
&t\mathbf{P}\big((\frac{X}{t}, \frac{Y}{t}) \in ([0,x] \times [0,y])^c\big) \to \frac{1}{x} + \frac{1}{y}, \quad x > 0, y > 0.
\end{aligned}
$$

The equations above verifies Theorem 2.2.2 and (Proposition 2.2.3).

We will now state and prove the result (Theorem 2.2.2) for a simple standard case which is a particularization of Theorem 2.2.1.

Proposition 2.2.4. *$(X, Y) \in \mathbb{R}_+^2$ is a bivariate random vector with the following properties:*

1. *(X, Y) has a density $f_{X,Y}(x, y)$.*
2. *Corresponding to regular variation in standardized form we assume,*

$$t^2 f_X(xt) \to x^{-2}, x > 1, \qquad t^2 f_Y(yt) \to y^{-2}, y > 1. \tag{2.2.30}$$

3. *The joint density satisfies*

$$t^3 f_{X,Y}(tx, ty) \to g_1(x, y) \in L_1(\mathbb{E}_\sqcap), \tag{2.2.31}$$

$$t^3 f_{X,Y}(tx, ty) \to g_2(x, y) \in L_1(\mathbb{E}_\sqsupset), \tag{2.2.32}$$

where $g_1(x, y), g_2(x, y) \ge 0$ are non-trivial, assumed 0 outside of $\mathbb{E}_\sqcap$ and $\mathbb{E}_\sqsupset$ respectively and we have

   (a) $v^2 g_1(u, v)$ is a probability density in v for each $u > 0$,

   (b) $u^2 g_2(u, v)$ is a probability density in u for each $v > 0$.

Then we have

$$tP\Big[\Big(\frac{X}{t},\frac{Y}{t}\Big)\in\cdot\ \Big]\xrightarrow{v}\mu(\cdot)\ \textit{in}\ \mathbb{M}_+\big(\mathbb{E}\big),$$

where

$$\mu(A)=\int\limits_{\{(u,v)\in A\}} g^*(u,v)dudv \qquad \textit{for } A\in\mathcal{E},$$

where $g^*(x,y)=g_1(x,y)\vee g_2(x,y)$.

Remark 2.2.4. Referring to Theorem 2.2.1, the conditions in (2.2.30) correspond to marginal convergences implicit in the Theorem 2.2.1, and conditions (2.2.31) and (2.2.32) correspond to conditions (2.2.3) and (2.2.4). Conditions (a) and (b) guarantee that we have proper conditional probability densities.

Proof. Clearly $g^*(\cdot)\in L_1(\mathbb{E})$. Also note that (2.2.31) and (2.2.32) imply that

$$g_1(x,y)=g_2(x,y)=g^*(x,y), \qquad \text{for } (x,y)\in\mathbb{E}_0.$$

Hence, for $(u,v)\in\mathbb{E}_0$,

$$\begin{aligned} tf_{\frac{X}{t},\frac{Y}{t}}(u,v)&=t^3f_{X,Y}(tu,tv)\\ &\to g^*(u,v), \qquad \text{as } t\to\infty, \end{aligned}$$

using (2.2.31) or (2.2.32). Now for $x>0,y>0$:

$$\begin{aligned} &t\mathbf{P}\Big[\Big(\frac{X}{t},\frac{Y}{t}\Big)\in\big([0,x]\times[0,y]\big)^c\Big]\\ &=t\mathbf{P}\Big(\frac{X}{t}>x\Big)+t\mathbf{P}\Big(\frac{Y}{t}>y\Big)-t\mathbf{P}\Big(\frac{X}{t}>x,\frac{Y}{t}>y\Big)\\ &=t\int_x^\infty f_{\frac{X}{t}}(u)du+t\int_y^\infty f_{\frac{Y}{t}}(v)dv-t\int_x^\infty\int_y^\infty f_{\frac{X}{t},\frac{Y}{t}}(u,v)dudv\\ &\to x^{-1}+y^{-1}-\int_x^\infty\int_y^\infty g^*(u,v)dudv, \qquad \text{as } t\to\infty, \end{aligned}$$

which is a consequence of Scheffé's Theorem. Also note that

$$x^{-1} = \int\limits_{x}^{\infty} u^{-2} du = \int\limits_{x}^{\infty} u^{-2} \Big(\int\limits_{0}^{\infty} u^2 g_2(u,v) dv \Big) du = \int\limits_{x}^{\infty} \int\limits_{0}^{\infty} g^*(u,v) dv du,$$

and similarly

$$y^{-1} = \int\limits_{0}^{\infty} \int\limits_{y}^{\infty} g^*(u,v) dv du.$$

Therefore as $t \to \infty$,

$$\begin{aligned} & t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in \big([0,x] \times [0,y]\big)^c\Big] \\ & \to \int\limits_{x}^{\infty} \int\limits_{0}^{\infty} g^*(u,v) dv du + \int\limits_{0}^{\infty} \int\limits_{y}^{\infty} g^*(u,v) dv du - \int\limits_{x}^{\infty} \int\limits_{y}^{\infty} g^*(u,v) du dv \\ & = \int\limits_{(u,v) \in ([0,x] \times [0,y])^c} g^*(u,v)(u,v) du dv = \mu\Big(\big([0,x] \times [0,y]\big)^c\Big). \end{aligned} \tag{2.2.33}$$

According to Lemma 6.1, Resnick [2007], proving 2.2.33 suffices for our proof.

□

2.2.4 Consistency in a d-dimensional set-up:

Suppose we have a d-dimensional vector $(X_1, X_2, \ldots, X_d)$ where a multivariate CEV model holds (with a definition similar to Definition 2.1.1) with a limit holding for whichever X_i we consider to be extreme. Then we can show that the distribution of $(X_1, X_2, \ldots, X_d)$ belongs to the domain of attraction of a d-dimensional extreme value distribution. The proof has not been provided here. We are still studying the multivariate CEV model and all results holding in a bivariate case does not naturally extend to the multivariate case.

2.3 Connecting regular variation on cones to the CEV model

We have seen in the previous sections that questions about the general conditional model are effectively analyzed by starting with standard regular variation on our special cones ($\mathbb{E}_{\sqcap}$ or $\mathbb{E}_{\sqcap}$). A pertinent question to ask here is, whether standardization of the conditional extreme value model is always possible. A partial answer has been provided in [Heffernan and Resnick, 2007, Section 2.4]. We consider this issue in more detail in this section. We start by making precise what we mean by *standardization*.

2.3.1 Standardization

Standardization is the process of marginally transforming a random vector $\boldsymbol{X}$ into a different vector $\boldsymbol{Z}^*$, $\boldsymbol{X} \mapsto \boldsymbol{Z}^*$, so that the distribution of $\boldsymbol{Z}^*$ is standard regularly varying on a cone $\mathbb{E}^*$. This means for some Radon measure $\mu^*(\cdot)$

$$t\mathbf{P}\Big[\frac{\boldsymbol{Z}^*}{t} \in \cdot\ \Big] \xrightarrow{v} \mu^*(\cdot) \text{ in } \mathbb{M}_+(\mathbb{E}^*).$$

In general, depending on the cone, this says one or more components of $\boldsymbol{Z}^*$ are asymptotically Pareto. For the classical multivariate extreme value theory case, each is asymptotically Pareto and then $\mathbb{E}^* = \mathbb{E} = [\mathbf{0}, \infty] \setminus \{\mathbf{0}\}$. The technique is used in classical multivariate extreme value theory to characterize multivariate domains of attraction and dates at least to de Haan and Resnick [1977]. See also [Resnick, 2008b, Chapter 5], Mikosch [2005, 2006], de Haan and Ferreira [2006], Resnick [2007].

Theoretical advantages of standardization:

- Standardization is analogous to the copula transformation but is better suited to studying limit relations [Klüppelberg and Resnick, 2008].
- In Cartesian coordinates, the limit measure has scaling property:

$$\mu^*(c\,\cdot\,) = c^{-1}\mu^*(\cdot), \quad c > 0.$$

- The scaling in Cartesian coordinates allows transformation to polar coordinates to yield a product measure: An angular measure exists allowing characterization of limits:

$$\mu^*\{\boldsymbol{x} : \|\boldsymbol{x}\| > r, \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} \in \Lambda\} = r^{-1}S(\Lambda),$$

  for Borel subsets Λ of the unit sphere in $\mathbb{E}^*$.

Note that for classical multivariate extreme value theory, S is a finite measure which we may take to be a probability measure without loss of generality. However, when $\mathbb{E}^* = \mathbb{E}_\sqcap$, S is NOT necessarily finite. This is because absence of the horizontal axis boundary in $\mathbb{E}_\sqcap$ implies the unit sphere is not compact.

Standardizing functions.

The most useful circumstance for standardization is discussed in the following definition.

Definition 2.3.1. *Suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_d)$ is a vector-valued random variable in $\mathbb{R}^d$ which satisfies:*

$$t\mathbf{P}\Big[\Big(\frac{X_1 - \beta_1(t)}{\alpha_1(t)}, \frac{X_2 - \beta_2(t)}{\alpha_2(t)}, \ldots, \frac{X_d - \beta_d(t)}{\alpha_d(t)}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+(\mathfrak{D}) \tag{2.3.1}$$

for some $\mathfrak{D} \subset \overline{\mathbb{R}}^d$, $\alpha_i(t) > 0, \beta_i(t) \in \mathbb{R}$, *for* $i = 1, \dots, d$. *Suppose we have* $\boldsymbol{f} = (f_1, \dots, f_d)$ *such that, for* $i = 1, \dots, d$:

(a) f_i : *Range of* $X_i \to (0, \infty)$,

(b) f_i *is monotone,*

(c) $\nexists K > 0$ *such that* $|f_i| \le K$.

Then $\boldsymbol{f}$ standardizes $\boldsymbol{X}$ *if* $\boldsymbol{Z}^* = \boldsymbol{f}(\boldsymbol{X})$ *satisfies*

$$
\begin{aligned}
t\mathbf{P}\Big[\frac{\boldsymbol{Z}^*}{t} \in \cdot\ \Big] &= t\mathbf{P}\Big[\Big(\frac{f_1(X_1)}{t}, \frac{f_2(X_2)}{t}, \dots, \frac{f_d(X_d)}{t}\Big) \in \cdot\ \Big] \\
&\xrightarrow{v} \mu^*(\cdot) \text{ in } \mathbb{M}_+(\mathbb{E}^*),
\end{aligned}
\tag{2.3.2}
$$

where $\mathbb{E}^*$ *is some cone in* $\overline{\mathbb{R}}_+^d$, *and* μ^* *is Radon. Call* $\boldsymbol{f}$ *the standardizing function and say* (2.3.2) *is the* standardization *of* (2.3.1).

For the conditional model defined in Definition 2.1.1 in Section 2.1.1 where F, the distribution of Y, satisfies $F \in D(G_1)$, we can always standardize Y with

$$
b^{\leftarrow}(\cdot) = \Big(\frac{1}{1-F}\Big)^{\leftarrow}(\cdot).
$$

We define $Y^* = b^{\leftarrow}(Y)$ to be the standardized version of Y and the standardizing function is $b^{\leftarrow}$. See Heffernan and Resnick [2007].

2.3.2 When can the CEV model be standardized?

Suppose Definition 2.1.1 holds; that is

$$
t\mathbf{P}\Big[\Big(\frac{X-\beta(t)}{\alpha(t)}, \frac{Y-b(t)}{a(t)}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+\big([-\infty,\infty] \times \overline{\mathbb{E}}^{(\gamma)}\big).
$$

Heffernan and Resnick [2007] show that standardization in the above equation is possible unless $(\psi_1, \psi_2) = (1, 0)$ which is equivalent to the limit measure being a product measure. We show that the converse is true too. So when the limit measure is not a product measure, we can always reduce to standard regular variation on a cone $\mathbb{E}_\sqcap$, and conversely we can think of the general conditional model as a transformation of standard regular variation on $\mathbb{E}_\sqcap$.

We begin with initial results about the impossibility of the limit measure being a product in the standardized convergence on $\mathbb{E}_\sqcap$, gradually leading to our final result in Proposition 2.3.3.

Lemma 2.3.1. *Suppose (X, Y) is standard regularly varying on the cone $\mathbb{E}_\sqcap$, such that,*

$$t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu(\cdot) \textit{ in } \mathbb{M}_+\big(\mathbb{E}_\sqcap\big) \tag{2.3.3}$$

for some non-null Radon measure $\mu(\cdot)$ on $\mathbb{E}_\sqcap$, satisfying the conditional non-degeneracy conditions as in (2.1.4) *and* (2.1.5)*. Then $\mu(\cdot)$ cannot be a product measure.*

Proof. If μ is a product measure we have

$$\mu\big([0, x] \times (y, \infty]\big) = G(x)y^{-1} \qquad \text{for } x \geq 0, y > 0 \tag{2.3.4}$$

for some finite distribution function G on $[0, \infty)$. Now (2.3.3) implies that μ is homogeneous of order -1, i.e.,

$$\mu(c\Lambda) = c^{-1}\mu(\Lambda), \qquad \forall c > 0, \tag{2.3.5}$$

where Λ is a Borel subset of $\mathbb{E}_\sqcap$. Therefore

$$\begin{aligned} \mu\big(c([0, x] \times (y, \infty])\big) &= \mu\big([0, cx] \times (cy, \infty]\big) \\ &= G(cx) \times \frac{1}{cy} \qquad\qquad \text{(using (2.3.4))} \\ &= c^{-1}G(cx)y^{-1}. \end{aligned}$$

Moreover, from (2.3.4) and (2.3.5) we have

$$\mu\big(c([0,x]\times(y,\infty])\big) = c^{-1}G(x)y^{-1}.$$

Therefore

$$G(cx) = G(x) \quad \forall c>0, x>0.$$

Hence for fixed $y \in \mathbb{E}^{(\gamma)}, c>0, x>0$,

$$\mu\big([0,cx]\times(y,\infty]\big) = G(cx)y^{-1} = G(x)y^{-1} = \mu\big([0,x]\times(y,\infty]\big).$$

Therefore μ becomes a degenerate distribution in x, contradicting our conditional non-degeneracy assumptions. Thus $\mu(\cdot)$ cannot be a product measure. □

Lemma 2.3.1 means that standard regular variation on $\mathbb{E}_\sqcap$ with a limit measure satisfying the conditional non-degeneracy conditions implies that the limit cannot be a product measure. Now suppose we have a generalized model as defined in Definition 2.1.1 and the limit measure is a product. We will show that we cannot *standardize* this to standard regular variation on some cone $\mathfrak{C} \subset \mathbb{E}$ ($\mathfrak{C} = \mathbb{E}_\sqcap$ for our case). Recall that when Definition 2.1.1 holds, we can always standardize Y so in the following we assume Y^* is standardized and only worry about the standardization of X.

Theorem 2.3.2. *Suppose $X \in \mathbb{R}, Y^* > 0$ are random variables, such that for functions $\alpha(\cdot) > 0, \beta(\cdot) \in \mathbb{R}$, we have*

$$t\mathbf{P}\Big[\Big(\frac{X-\beta(t)}{\alpha(t)}, \frac{Y^*}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} G \times \nu_1(\cdot) \text{ in } \mathbb{M}_+\big([-\infty,\infty]\times(0,\infty]\big), \tag{2.3.6}$$

as $t \to \infty$, where $\nu_1(x,\infty] = x^{-1}$, $x > 0$, and G is some finite, non-degenerate distribution on $\mathbb{R}$. Then there does not exist a standardizing function, $f(\cdot)$: Range of $X \to (0,\infty)$, in the sense of Definition 2.3.1, such that

$$t\mathbf{P}\Big[\Big(\frac{f(X)}{t}, \frac{Y^*}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}_\sqcap\big), \tag{2.3.7}$$

where μ satisfies the conditional non-degeneracy conditions.

Proof. Note that Y^* is already standardized here. Suppose there exists a standardizing function $f(\cdot)$ such that (2.3.7) holds. Without loss of generality assume $f(\cdot)$ to be non-decreasing. This implies that for $\mu-$ continuity points (x,y) we have,

$$t\mathbf{P}\Big[\frac{f(X)}{t} \le x, \frac{Y^*}{t} > y\Big] \to \mu\big((-\infty,x] \times (y,\infty]\big) \qquad (t \to \infty)$$

which is equivalent to

$$t\mathbf{P}\Big(\frac{X-\beta(t)}{\alpha(t)} \le \frac{f^{\leftarrow}(xt)-\beta(t)}{\alpha(t)}, \frac{Y^*}{t} > y\Big] \to \mu\big((-\infty,x] \times (y,\infty]\big), \qquad (t \to \infty). \tag{2.3.8}$$

Since $\mu((-\infty,x] \times (y,\infty]) < \infty$ and is non-degenerate in x, we have as $t \to \infty$ that

$$\frac{f^{\leftarrow}(xt)-\beta(t)}{\alpha(t)} \to h(x) \tag{2.3.9}$$

for some non-decreasing function $h(\cdot)$ which has at least two points of increase. Thus (2.3.8) and (2.3.9) imply that

$$\mu((-\infty,x] \times (y,\infty]) = G(h(x)) \times y^{-1}.$$

Hence $\mu(\cdot)$ turns out to be a product measure which by Lemma 2.3.1 is not possible. $\square$

The final result of this section shows that one can transform from the conditional extreme value model (like Definition 2.1.1) to the standard model (like equation (2.3.3)) and vice-versa if and only if the the limit measure in the generalized model is not a product measure.

Proposition 2.3.3. *We have two parts to this proposition.*

1. *Suppose we have the conditional extreme value model of Definition 2.1.1; i.e., we have a random vector* $(X, Y) \in \mathbb{R}^2$*, and there exists functions* $a(t) > 0, b(t) \in \mathbb{R}, \alpha(t) > 0, \beta \in \mathbb{R}$*, such that for* $\gamma \in \mathbb{R}$*,*

   $$t\mathbf{P}\Big[\Big(\frac{X-\beta(t)}{\alpha(t)}, \frac{Y-b(t)}{a(t)}\Big)\cdot\Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+\big([-\infty,\infty] \times \overline{\mathbb{E}}^{(\gamma)}\big),$$

   along with the conditional non-degeneracy conditions (2.1.4) *and* (2.1.5)*. Hence equation* (2.1.10) *holds; i.e.,*

   $$\lim_{t\to\infty} \frac{\alpha(tc)}{\alpha(t)} = \psi_1(c), \qquad \lim_{t\to\infty} \frac{\beta(tc)-\beta(t)}{\alpha(t)} = \psi_2(c). \tag{2.3.10}$$

   If $(\psi_1, \psi_2) \neq (1, 0)$*, then there exists a standardization function* $\boldsymbol{f} = (f_1, f_2)$ *such that* $(X^*, Y^*) = (f_1(X), f_2(Y))$ *is standard regularly varying on* $\mathbb{E}_\sqcap$*; that is*

   $$t\mathbf{P}\Big[\Big(\frac{f_1(X)}{t}, \frac{f_2(Y)}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu^{**}(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}_\sqcap\big),$$

   where μ^{**} *is a non-null Radon measure satisfying the conditional non-degeneracy conditions.*

2. *Conversely, suppose we have a bivariate random vector* $(X^*, Y^*) \in \mathbb{R}_+^2$ *satisfying*

   $$t\mathbf{P}\Big[\Big(\frac{X^*}{t}, \frac{Y^*}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu^{**}(\cdot) \text{ in } \mathbb{M}_+\big(\mathbb{E}_\sqcap\big),$$

   where μ^{**} *is a non-null Radon measure, satisfying the conditional non-degeneracy conditions. Consider functions* $\alpha(\cdot) > 0$*,* $\beta(\cdot) \in \mathbb{R}$ *such that* *equation* (2.3.10) *holds with* $(\psi_1, \psi_2) \neq (1, 0)$. *Then there exist functions* $a(\cdot) > 0,\ b(\cdot) \in \mathbb{R}$ *satisfying* (2.1.2) *and* $\lambda(\cdot) \in \mathbb{R}$, $\gamma \in \mathbb{R}$ *such that*

$$tP\Big[\Big(\frac{\lambda(X^*) - \beta(t)}{\alpha(t)}, \frac{b(Y^*) - b(t)}{a(t)}\Big) \in \cdot\ \Big] \xrightarrow{v} \tilde{\mu}(\cdot) \tag{2.3.11}$$

in $\mathbb{M}_+\big([-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)}\big)$ *where* $\tilde{\mu}$ *is a non-null Radon measure satisfying the conditional non-degeneracy conditions and* $b(Y^*) \in D(G_\gamma)$.

Remark 2.3.1. The class of limit measures in Definition 2.1.1 which are not product measures can thus be considered to be obtained from standard regular variation on $\mathbb{E}_\sqcap$ after appropriate marginal transformations.

Proof. (1) This part has been dealt with in [Heffernan and Resnick, 2007, Section 2.4].

(2) First we simplify the problem. Note that, for (x, y) a continuity point of $\mu(\cdot)$,

$$tP\Big[\frac{\lambda(X^*) - \beta(t)}{\alpha(t)} \leq x, \frac{b(Y^*) - b(t)}{a(t)} > y\Big] \to \tilde{\mu}\big([-\infty, x] \times (y, \infty]\big) \qquad (t \to \infty)$$

is equivalent to

$$\begin{aligned} tP\Big(\frac{\lambda(X^*) - \beta(t)}{\alpha(t)} \leq x, \frac{Y^*}{t} > y\Big) &\to \tilde{\mu}\big([-\infty, x] \times (h(y), \infty]\big) \qquad (t \to \infty) \\ &=: \mu^*\big([-\infty, x] \times (y, \infty]\big) \end{aligned} \tag{2.3.12}$$

where

$$h(y) = \begin{cases} (1 + \gamma y)^{\frac{1}{\gamma}} & \gamma \neq 0 \\ e^y & \gamma = 0. \end{cases} \tag{2.3.13}$$

Hence (2.3.11) is equivalent to

$$tP\Big[\Big(\frac{\lambda(X^*) - \beta(t)}{\alpha(t)}, \frac{Y^*}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu^*(\cdot)$$

and μ^* is a non-null Radon measure on $[-\infty, \infty] \times \bar{\mathbb{E}}^{(\gamma)}$ satisfying the conditional non-degeneracy conditions. Hence our proof will show the existence of $\lambda(\cdot)$ satisfying (2.3.12). Now note that equation (2.3.10) implies that $\alpha(\cdot) \in RV_\rho$ for some $\rho \in \mathbb{R}$ and $\psi_1(x) = x^\rho$ (see [Resnick, 2008b, page 14]). The function $\psi_2(\cdot)$ may be identically equal to 0, or

$$\psi_2(x) = \begin{cases} k(x^\rho - 1)/\rho, & \text{if } \rho \neq 0, x > 0 \\ k \log x & \text{if } \rho = 0, x > 0 \end{cases} \tag{2.3.14}$$

for $k \neq 0$ [de Haan and Ferreira, 2006, page 373]. We have assumed that $(\psi_1, \psi_2) \neq (1, 0)$. We will consider three cases: $\rho > 0, \rho = 0, \rho < 0$.

Case 1 : $\rho > 0$.

1. Suppose $\psi_2 \equiv 0$.

   Since $\alpha(\cdot) \in RV_\rho$, there exists $\tilde{\alpha}(\cdot) \in RV_\rho$ which is ultimately differentiable and strictly increasing and $\alpha \sim \tilde{\alpha}$ [de Haan and Ferreira, 2006, page 366]. Thus $\tilde{\alpha}^{\leftarrow}$ exists. Additionally, we have from [Bingham et al., 1987, Theorem 3.1.12(a)], that $\beta(t)/\alpha(t) \to 0$. Hence we have for $x > 0$, as $t \to \infty$,

   $$\frac{\tilde{\alpha}(tx) + \beta(t)}{\alpha(t)} = \frac{\tilde{\alpha}(tx)}{\tilde{\alpha}(t)} \cdot \frac{\tilde{\alpha}(t)}{\alpha(t)} + \frac{\beta(t)}{\alpha(t)} \to x^\rho,$$

   and inverting we get for $z > 0$

   $$\frac{\tilde{\alpha}^{\leftarrow}\big(\alpha(t)z + \beta(t)\big)}{t} \to z^{1/\rho} \qquad (t \to \infty).$$

   Thus we have,

   $$\begin{aligned} t\mathbf{P}\Big[\frac{\tilde{\alpha}(X^*) - \beta(t)}{\alpha(t)} \leq x, \frac{Y^*}{t} > y\Big] &= t\mathbf{P}\Big[\frac{X^*}{t} \leq \frac{\tilde{\alpha}^{\leftarrow}(\alpha(t)x + \beta(t))}{t}, \frac{Y^*}{t} > y\Big] \\ &\to \mu^{**}\big([0, x^{1/\rho}] \times (y, \infty]\big). \end{aligned}$$

   Set $\lambda(\cdot) = \tilde{\alpha}(\cdot)$ and this defines $\tilde{\mu}$.

2. Now suppose $\psi_2 \neq 0$.

   Therefore

$$\psi_2(x) = \lim_{t\to\infty} \frac{\beta(tx) - \beta(t)}{\alpha(t)} = k(x^\rho - 1)/\rho;$$

   that is, $\beta(\cdot) \in RV_\rho$ and $k > 0$. There exists $\tilde{\beta}$ which is ultimately differentiable and strictly increasing and $\tilde{\beta} \sim \beta$ [de Haan and Ferreira, 2006, page 366]. Thus $\tilde{\beta}^{\leftarrow}$ exists. Then we have for $x > 0$, as $t \to \infty$,

$$\begin{aligned}
\frac{\tilde{\beta}(tx) - \beta(t)}{\alpha(t)} &= \frac{\tilde{\beta}(tx) - \beta(tx)}{\alpha(t)} + \frac{\beta(tx) - \beta(t)}{\alpha(t)} \\
&= \frac{\tilde{\beta}(tx) - \beta(tx)}{\beta(tx)} \frac{\beta(tx)}{\alpha(tx)} \frac{\alpha(tx)}{\alpha(t)} + \frac{\beta(tx) - \beta(t)}{\alpha(t)} \\
&\to (1-1) \cdot \frac{1}{\rho} \cdot x^\rho + k\frac{x^\rho - 1}{\rho} = k\frac{x^\rho - 1}{\rho}.
\end{aligned}$$

   Inverting, we get as $t \to \infty$,

$$\frac{\tilde{\beta}^{\leftarrow}\big(\alpha(t)x + \beta(t)\big)}{t} \overset{t\to\infty}{\to} (1 + \frac{\rho x}{k})^{1/\rho}.$$

   Thus we have,

$$\begin{aligned}
t\mathbf{P}\Big[\frac{\tilde{\beta}(X^*) - \beta(t)}{\alpha(t)} \leq x, \frac{Y^*}{t} > y\Big] &= t\mathbf{P}\Big[\frac{X^*}{t} \leq \frac{\tilde{\beta}^{\leftarrow}(\alpha(t)x + \beta(t))}{t}, \frac{Y^*}{t} > y\Big] \\
&\to \mu^{**}\big([0, (1 + \frac{\rho x}{k})^{1/\rho}] \times (y, \infty]\big).
\end{aligned}$$

   Here we can set $\lambda(\cdot) = \tilde{\beta}(\cdot)$ and this defines $\tilde{\mu}$.

Case 2: $\rho = 0$.

We have $\psi_1(x) = 1, \psi_2(x) = k \log x$ for $x > 0$ and some $k \in \mathbb{R}$. By assumption, $(\psi_1, \psi_2) \neq (1, 0)$ and hence $k \neq 0$. First assume that $k > 0$, which means $\beta \in \Pi_+(\alpha)$. From property (2) for π-varying functions (Section A.1.1), there exists $\tilde{\beta}(\cdot)$ which is continuous, strictly increasing and $\beta - \tilde{\beta} = o(\alpha)$. If $\beta(\infty) = \tilde{\beta}(\infty) =$

∞, we have for $x > 0$,

$$\frac{\tilde{\beta}(tx) - \beta(t)}{\alpha(t)} = \frac{\tilde{\beta}(tx) - \beta(tx)}{\alpha(tx)} \frac{\alpha(tx)}{\alpha(t)} + \frac{\beta(tx) - \beta(t)}{\alpha(t)}$$

$$\to 0 + k \log x,$$

and inverting, we get for $z \in \mathbb{R}$, as $t \to \infty$,

$$\frac{\tilde{\beta}^{\leftarrow}\big(\alpha(t)z + \beta(t)\big)}{t} \to e^{z/k}.$$

Thus we have,

$$t\mathbf{P}\Big(\frac{\tilde{\beta}(X^*) - \beta(t)}{\alpha(t)} \leq x, \frac{Y^*}{t} > y\Big) = t\mathbf{P}\Big(\frac{X^*}{t} \leq \frac{\tilde{\beta}^{\leftarrow}(\alpha(t)x + \beta(t))}{t}, \frac{Y^*}{t} > y\Big)$$

$$\to \mu\big([0, e^{k/x}] \times (y, \infty]\big).$$

If $\beta(\infty) = \tilde{\beta}(\infty) = B < \infty$, define

$$\beta^*(t) = \frac{1}{B - \tilde{\beta}(t)}, \qquad \alpha^*(t) = \frac{\alpha(t)}{(B - \tilde{\beta}(t))^2}$$

and from property 3(b), for π-varying functions in Section A.1.1, we have that $\beta^* \in \Pi_+(\alpha^*)$, $\beta^*(t) \to \infty$ and $\frac{B - \tilde{\beta}(t)}{\alpha(t)} \stackrel{t\to\infty}{\to} \infty$. Hence we have reduced to the previous case which implies,

$$t\mathbf{P}\Big(\frac{\beta^*(X^*) - \beta^*(t)}{\alpha^*(t)} \leq x, \frac{Y^*}{t} > y\Big) \to \mu\big([0, e^{k/x}] \times (y, \infty]\big).$$

This is equivalent to

$$t\mathbf{P}\Big(\frac{\tilde{\beta}(X^*) - \tilde{\beta}(t)}{\alpha(t)} \leq \frac{x}{1 + \frac{\alpha(t)x}{B - \tilde{\beta}(t)}}, \frac{Y^*}{t} > y\Big) \to \mu\big([0, e^{k/x}] \times (y, \infty]\big)$$

and since $\frac{B - \tilde{\beta}(t)}{\alpha(t)} \stackrel{t\to\infty}{\to} \infty$ implies $\frac{\alpha(t)}{B - \tilde{\beta}(t)} \stackrel{t\to\infty}{\to} 0$, we can write

$$t\mathbf{P}\Big(\frac{\tilde{\beta}(X^*) - \tilde{\beta}(t)}{\alpha(t)} \leq x, \frac{Y^*}{t} > y\Big) \to \mu\big([0, e^{k/x}] \times (y, \infty]\big)$$

which implies that

$$t\mathbf{P}\Big(\frac{\tilde{\beta}(X^*)-\beta(t)}{\alpha(t)} \le x, \frac{Y^*}{t} > y\Big) \to \mu\big([0, e^{k/x}] \times (y, \infty]\big) \quad \text{since, } \beta - \tilde{\beta} = o(\alpha)$$

and we have produced the required transformation $\lambda(\cdot) = \tilde{\beta}(\cdot)$.

The case for which $k < 0$; i.e., $\beta \in \Pi_-(\alpha)$ can be proved similarly.

Case 3 : $\rho < 0$. This case is similar to the case for $\rho > 0$.

1. Suppose $\psi_2 \equiv 0$.

   Since $\alpha(\cdot) \in RV_\rho$, there exists $\tilde{\alpha}(\cdot) \in RV_\rho$ which is ultimately differentiable and strictly decreasing and $\alpha \sim \tilde{\alpha}$ [de Haan and Ferreira, 2006, page 366]. Thus $\tilde{\alpha}^{\leftarrow}$ exists. Additionally, we have from Bingham et al. [1987], Theorem 3.1.10(a),(c) that $\beta(\infty) := \lim_{t\to\infty} \beta(t)$ exists and is finite, and $(\beta(\infty) - \beta(t))/\alpha(t) \to 0$.

$$\frac{\tilde{\alpha}(tx) + \beta(t) - \beta(\infty)}{\alpha(t)} = \frac{\tilde{\alpha}(tx)}{\tilde{\alpha}(t)} \cdot \frac{\tilde{\alpha}(t)}{\alpha(t)} + \frac{\beta(t) - \beta(\infty)}{\alpha(t)} \overset{t\to\infty}{\to} x^\rho,$$

   inverting which we get, for $z > 0$, as $t \to \infty$,

$$\frac{\tilde{\alpha}^{\leftarrow}\big(\alpha(t)z + \beta(\infty) - \beta(t)\big)}{t} \to z^{1/\rho}.$$

   Thus we have, now taking $x < 0$,

$$\begin{aligned} &t\mathbf{P}\Big(\frac{\beta(\infty) - \tilde{\alpha}(X^*) - \beta(t)}{\alpha(t)} \le x, \frac{Y^*}{t} > y\Big) \\ =&t\mathbf{P}\Big(\frac{X^*}{t} \ge \frac{\tilde{\alpha}^{\leftarrow}(-\alpha(t)x + \beta(\infty) - \beta(t))}{t}, \frac{Y^*}{t} > y\Big) \\ \to&\mu\big([(-x)^{1/\rho}, \infty] \times (y, \infty]\big) =: \tilde{\mu}\big([-\infty, x] \times (y, \infty]\big). \end{aligned}$$

   Therefore we can set $\lambda(\cdot) = \beta(\infty) - \tilde{\alpha}(\cdot)$.

2. Now suppose $\psi_2 \neq 0$.

   Therefore $\psi_2(x) = \lim_{t\to\infty} \frac{\beta(tx)-\beta(t)}{\alpha(t)} = k(x^\rho - 1)/\rho$; i.e., $\beta(\cdot) \in RV_\rho$ and $k < 0$. There exists $\tilde{\beta} \in RV_\rho$ which is ultimately differentiable and strictly decreasing and $\tilde{\beta} \sim \beta$ [de Haan and Ferreira, 2006, page 366]. Thus $\tilde{\beta}^{\leftarrow}$ exists. We also have $\beta(\infty) := \lim_{t\to\infty} \beta(t)$ exists and is finite, and $(\beta(\infty) - \beta(t))/\alpha(t) \to \frac{k}{|\rho|}$ [de Haan and Ferreira, 2006, page 373]. Then we have for $x > 0$, as $t \to \infty$,

$$\frac{\tilde{\beta}(tx) - \beta(t)}{\alpha(t)} \to k\frac{x^\rho - 1}{\rho},$$

   inverting which we get, as $t \to \infty$,

$$\frac{\tilde{\beta}^{\leftarrow}\big(\alpha(t)x + \beta(t)\big)}{t} \to (1 + \frac{\rho x}{k})^{1/\rho}.$$

   Thus we have,

$$\begin{aligned} t\mathbf{P}\Big(\frac{\tilde{\beta}(X^*) - \beta(t)}{\alpha(t)} \le x, \frac{Y^*}{t} > y\Big) &= t\mathbf{P}\Big(\frac{X^*}{t} \le \frac{\tilde{\beta}^{\leftarrow}(\alpha(t)x + \beta(t))}{t}, \frac{Y^*}{t} > y\Big) \\ &\to \mu\big([0, (1 + \frac{\rho x}{k})^{1/\rho}] \times (y, \infty]\big). \end{aligned}$$

   Here we can set $\lambda(\cdot) = \tilde{\beta}(\cdot)$.

Hence the result. $\square$

Remark 2.3.2. Suppose that we have

$$t\mathbf{P}\Big[\Big(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y^*}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu^*(\cdot) = H \times \nu_1(\cdot),$$

in $\mathbb{M}_+([-\infty, \infty] \times (0, \infty])$. We are assuming $\mu^*(\cdot)$ is a product measure. Let

$$X^* = \frac{(X - \beta(Y^*))Y^*}{\alpha(Y^*)}.$$

Then for continuity points (x, y) of the limit

$$t\mathbf{P}\Big(\frac{X^*}{t} \le x, \frac{Y^*}{t} > y\Big) \to \int_0^{1/y} H(xv)dv$$

in $\mathbb{M}_+([-\infty, \infty] \times (0, \infty])$. It is easy to check that the limit measure is homogeneous of order -1. Thus, a standardization of (X, Y^*) exists even when we have a limit measure which is a product. Note this standardization is not in the sense of Definition 2.3.1, and it represents a change of co-ordinate system which is more complex than just a marginal transformation.

2.3.3 A characterization of regular variation on $\mathbb{E}_\sqcap$

Standard regular variation on $\mathbb{E}$ was characterized by de Haan [1978] in terms of one dimensional regular variation of max linear combinations and Resnick [2002] provides a characterization of hidden regular variation in $\mathbb{E}$ and $\mathbb{E}_0$ in terms of max and min linear combinations of the random vector respectively. We provide a result in the same spirit for regular variation on $\mathbb{E}_\sqcap$.

Proposition 2.3.4. *Suppose $(X, Y) \in \mathbb{R}^2$ is a random vector and $\mathbf{P}(X = 0) = 0$. Then the following are equivalent:*

1. *(X, Y) is standard multivariate regularly varying on $\mathbb{E}_\sqcap$ with a limit measure satisfying the non-degeneracy conditions* (2.1.4) *and* (2.1.5).
2. *For all $a \in (0, \infty]$ we have*

$$\lim_{t\to\infty} tP\Big(\frac{\min(aX, Y)}{t} > y\Big) = c(a)y^{-1}, \qquad y > 0,$$

for some non-constant, non-decreasing function $c : (0, \infty] \to (0, \infty)$.

Proof. Since $\mathbf{P}(X = 0) = 0$ we have $\mathbf{P}(X > 0) = 1$.

(2) ⇒ (1): Assume that

$$\lim_{t\to\infty} tP\Big(\frac{\min(aX, Y)}{t} > y\Big) = c(a)y^{-1}, \qquad y > 0,$$

for some function $c : (0, \infty] \to (0, \infty)$. Then for $x \geq 0, y > 0$,

$$\begin{aligned} t\mathbf{P}\Big(\frac{X}{t} \leq x, \frac{Y}{t} > y\Big) &= t\mathbf{P}\Big(\frac{Y}{t} > y\Big) - t\mathbf{P}\Big(\frac{X}{t} > x, \frac{Y}{t} > y\Big) \\ &= t\mathbf{P}\Big(X > 0, \frac{Y}{t} > y\Big) - t\mathbf{P}\Big(\frac{(y/x)X}{t} > y, \frac{Y}{t} > y\Big) \\ &= t\mathbf{P}\Big(\frac{\min(a_1X, Y)}{t} > y\Big) - t\mathbf{P}\Big(\frac{\min((y/x)X, Y)}{t} > y\Big) \end{aligned}$$

where $a_1 = \infty$ and the above quantity

$$\to c(\infty)y^{-1} - c(y/x)y^{-1} =: \nu([0, x] \times (y, \infty]).$$

Since $c(\cdot)$ is non-decreasing and non-constant, ν is a non-null Radon measure on $\mathbb{E}_\sqcap$ and we have our result. The non-degeneracy of ν follows from the fact that $c(\cdot)$ is a non-constant function.

(1) ⇒ (2): Now assume that (X, Y) is standard multivariate regularly varying on $\mathbb{E}_\sqcap$. Hence there exists a non-degenerate Radon measure ν on $\mathbb{E}_\sqcap$ such that

$$\lim_{t\to\infty} t\mathbf{P}\Big(\frac{X}{t} \leq x, \frac{Y}{t} > y\Big) = \nu([0, x] \times (y, \infty]),$$

and for any $a \in (0, \infty]$

$$\begin{aligned} t\mathbf{P}\Big(\frac{\min(aX, Y)}{t} > y\Big) &= t\mathbf{P}\Big(\frac{X}{t} > \frac{y}{a}, \frac{Y}{t} > y\Big) \\ &= \nu((\frac{y}{a}, \infty] \times (y, \infty]) \\ &= y^{-1}\nu((\frac{1}{a}, \infty] \times (1, \infty]) =: c(a)y^{-1}, \end{aligned}$$

by defining $c(a) = \nu((\frac{1}{a}, \infty] \times (1, \infty])$ and using the homogeneity property (2.3.16). Note that the conditional non-degeneracy of ν implies that c is non-constant and non-decreasing.

Hence the result. $\square$

Remark 2.3.3. The condition $\mathbf{P}(X = 0)$ can be removed if we assume Y to be heavy-tailed with exponent $\alpha = 1$, i.e., as $t \to \infty$, $t\mathbf{P}(\frac{Y}{t} > y) \to y^{-1}$.

2.3.4 Polar co-ordinates

Proposition 2.3.3 shows that when the limit measure is not a product measure, we can transform (X, Y) to (X^*, Y^*) such that

$$\mathbf{P}\Big[\Big(\frac{X^*}{t}, \frac{Y^*}{t}\Big) \in \cdot\ \Big] \xrightarrow{v} \mu^{**}(\cdot) \text{ in } \mathbb{M}_+(\mathbb{E}_\sqcap) \tag{2.3.15}$$

Hence from Remark 2.2.2 we have that μ^{**} is homogeneous of order -1:

$$\mu^{**}(cB) = c^{-1}\mu^{**}(B), \quad \forall c > 0, B \in \mathcal{E}_\sqcap. \tag{2.3.16}$$

Hence μ^{**} has a spectral form. Further discussion on the spectral form is available in [Heffernan and Resnick, 2007, Section 3.2]. We provide a few facts here. For convenience let us take the norm

$$\|(x, y)\| = |x| + |y|, \quad (x, y) \in \mathbb{R}^2,$$

although any other norm would work too. Now, the standard argument using homogeneity [Resnick, 2008b, Chapter 5] yields for $r > 0$ and Λ a Borel subset of $[0, 1)$,

$$\begin{aligned} &\mu^{**}\Big\{(x, y) \in [0, \infty] \times (0, \infty] : x + y > r, \frac{x}{x + y} \in \Lambda\Big\} \\ &= r^{-1}\mu^{**}\Big\{(x, y) \in [0, \infty] \times (0, \infty] : x + y > 1, \frac{x}{x + y} \in \Lambda\Big\} =: r^{-1}S(\Lambda). \end{aligned} \tag{2.3.17}$$

where S is a Radon measure on $[0,1)$. Note that from (2.3.17), we can calculate for $x>0, y>0$,

$$\mu^{**}\big([0,x]\times(y,\infty]\big) = y^{-1}\int_0^{x/(x+y)}(1-w)S(dw) - x^{-1}\int_0^{x/(x+y)} wS(dw). \tag{2.3.18}$$

S need not be a finite measure on $[0,1)$ but to guarantee that

$$H^{**}(x) := \mu^{**}\big([0,x]\times(1,\infty]\big) \tag{2.3.19}$$

is a probability measure, we can see by taking $x\to\infty$ in (2.3.18) that we need

$$\int_0^1 (1-w)S(dw) = 1. \tag{2.3.20}$$

Conclusion: The class of limits μ^{**} in (2.3.15) or conditional limits

$$H^{**}(x) = \lim_{t\to\infty}\mathbf{P}\Big[\frac{X^*}{t}\le x|Y>t\Big]$$

is indexed by Radon measures S on $[0,1)$ satisfying the integrability condition (2.3.20).

Example 2.3.1 (Finite angular measure). Suppose S is uniform on $[0,1)$, $S(dw)=2dw$, so that equation (2.3.20) is satisfied. Then we have

$$\mu^{**}([0,x]\times(y,\infty]) = \frac{x}{y(x+y)}. \tag{2.3.21}$$

Putting $y=1$ we get that

$$H^{**}(x) = 1-\frac{1}{1+x}, \quad x>0,$$

which is a Pareto distribution.

Example 2.3.2 (Infinite angular measure). Now suppose S has angular measure $S(dw)=\frac{1}{1-w}dw$. This also satisfies equation (2.3.20). Now we have

$$\mu^{**}([0,x]\times(y,\infty]) = \frac{1}{y}+\frac{1}{x}\log(1-\frac{x}{x+y}). \tag{2.3.22}$$

Putting $y = 1$ we get that

$$H^{**}(x) = 1 - \frac{1}{x}\log(1+x), \quad x > 0.$$

Here H^{**} is continuously increasing, $\lim_{x\downarrow 0} H^{**}(x) = 0$ and $\lim_{x\uparrow\infty} H^{**}(x) = 1$, and hence $H^{**}(\cdot)$ is a valid distribution function. Note that S has infinite angular measure. One way to get infinite angular measures satisfying (2.3.20) is to take $S(dw) = \frac{1}{1-w}F(dw)$ for probability measures $F(\cdot)$ on $[0,1)$.

2.4 Extending the CEV model to a multivariate extreme value model

Observe that the CEV model assumes the existence of a vague limit in a subset of the Euclidean space which is smaller than that for classical MEVT. So it is natural to ask when can we extend a CEV model to a MEVT model. We answer this question in the current section. Clearly, any extension of the CEV model to MEVT will require X to also have a distribution in a domain of attraction. The first Proposition provides a sufficient condition for such an extension.

Proposition 2.4.1. *Suppose we have $(X,Y) \in \mathbb{R}^2$ and non-negative functions $\alpha(\cdot)$, $a(\cdot)$ and real functions $\beta(\cdot)$, $b(\cdot)$ such that*

$$t\mathbf{P}\Big[\Big(\frac{X-\beta(t)}{\alpha(t)}, \frac{Y-b(t)}{a(t)}\Big) \in \cdot\,\Big] \xrightarrow{v} \mu(\cdot) \text{ in } \mathbb{M}_+\big([-\infty,\infty]\times\overline{\mathbb{E}}^{(\gamma)}\big),$$

for some $\gamma \in \mathbb{R}$ where μ satisfies the appropriate conditional non-degeneracy conditions corresponding to (2.1.4)–(2.1.6). *Also assume that $X \in D(G_\lambda)$ for some $\lambda \in \mathbb{R}$; i.e., there exists functions $\chi(t) > 0, \phi(t) \in \mathbb{R}$ such that for continuity points $x \in \mathbb{E}^{(\lambda)}$ of the limit G_λ we have*

$$t\mathbf{P}\Big(\frac{X-\phi(t)}{\chi(t)} > x\Big) \to (1+\lambda x)^{-1/\lambda}, \quad 1+\lambda x > 0.$$

If $\lim_{t\to\infty} \chi(t)/\alpha(t)$ *exists and is in* $(0,\infty]$ *then* (X,Y) *is in the domain of attraction of a multivariate extreme value distribution on* $\mathbb{E}^{(\lambda,\gamma)}$*; that is,*

$$t\mathbf{P}\Big[\Big(\frac{X-\phi(t)}{\chi(t)}, \frac{Y-b(t)}{a(t)}\Big) \in \cdot\ \Big] \xrightarrow{v} (\mu \diamond \nu)(\cdot) \textit{ in } \mathbb{M}_+\big(\mathbb{E}^{(\lambda,\gamma)}\big)$$

where $(\mu \diamond \nu)(\cdot)$ *is a Radon measure on* $\mathbb{E}^{(\lambda,\gamma)}$.

Proof. The proof is a consequence of cases 1 and case 2 of Theorem 2.2.2. □

In the next result we characterize extension of CEV model to MEVT in terms of polar co-ordinates. Assume that (X,Y) can be standardized to (X^*,Y^*) which is regularly varying on $\mathbb{E}_\sqcap$. The following result provides a sufficient condition for an extension of regular variation on $\mathbb{E}_\sqcap$ to an asymptotically tail equivalent regular variation on $\mathbb{E}$. A short discussion on multivariate tail equivalence is provided in the appendix in Section A.1.2.

Proposition 2.4.2. *Suppose* $(X,Y) \in \mathbb{R}^2$ *is standard regularly varying on the cone* $\mathbb{E}_\sqcap$ *with limit measure* $\nu_\sqcap$ *and angular measure* $S_\sqcap$ *on* $[0,1)$. *Then the following are equivalent.*

1. $S_\sqcap$ *is finite on* $[0,1)$.
2. *There exists a random vector* (X^*,Y^*) *defined on* $\mathbb{E}$ *such that*

$$(X^*,Y^*) \overset{te(\mathbb{E}_\sqcap)}{\sim} (X,Y)$$

and (X^*,Y^*) *is multivariate regularly varying on* $\mathbb{E}$ *with limit measure* ν *such that* $\nu|_{\mathbb{E}_\sqcap} = \nu_\sqcap$.

Proof. Consider each implication separately.

$(1) \Rightarrow (2)$: Define the polar coordinate transformation $(R, \Theta) = (X + Y, \frac{X}{X+Y})$. From Section 2.3.4 we have any $r > 0$, and Λ a Borel subset of $[0, 1)$, as $t \to \infty$,

$$t\mathbf{P}\Big[\frac{R}{t} > r, \Theta \in \Lambda\Big] \to r^{-1}S_{\sqcap}(\Lambda).$$

Note equation (2.3.17) implies that the right side in the previous line is also equal to

$$\nu_{\sqcap}\{(x,y) \in \mathbb{E}_{\sqcap} : x + y > r, \frac{x}{x+y} \in \Lambda\}.$$

Since $S_{\sqcap}$ is finite on [0,1), the distribution of Θ is finite on $[0, 1)$. Assume $S[0, 1) = 1$ so that it is a probability measure and extend the measure $S_{\sqcap}$ to $[0, 1]$ by putting $S_{\sqcap}(\{1\}) = 0$. Let us define R_0 and Θ_0 such that they are independent, Θ_0 has distribution given by the extended $S_{\sqcap}$ on [0,1] and R_0 has the standard Pareto distribution. Define

$$(X^*, Y^*) = (R_0\Theta_0, R_0(1 - \Theta_0)).$$

Clearly (X^*, Y^*) is regularly varying on $\mathbb{E}$, now with the standard scaling and limit measure ν where $\nu|_{\mathbb{E}_{\sqcap}} = \nu_{\sqcap}$.

$(2) \Rightarrow (1)$: Referring to (2.3.17) note that

$$S_{\sqcap}\big([0, 1)\big) = \nu_{\sqcap}\{(x, y) \in \mathbb{E}_{\sqcap} : x + y > 1\}.$$

Since (X^*, Y^*) is regularly varying on $\mathbb{E}$, we have

$$\begin{aligned} t\mathbf{P}(\frac{R}{t} > 1) &= t\mathbf{P}(\frac{X+Y}{t} > 1) \\ &\to \nu\{(x,y) \in \mathbb{E}_{\sqcap} : x + y > 1\} < \infty. \end{aligned}$$

But

$$\nu\{(x, y) \in \mathbb{E}_{\sqcap} : x + y > 1\} = \nu_{\sqcap}\{(x, y) \in \mathbb{E}_{\sqcap} : x + y > 1\} = S_{\sqcap}\big([0, 1)\big).$$

Hence $S_{\sqcap}$ is finite on $[0, 1)$.

Thus we have shown both the implications. $\square$

2.5 Examples

In this section we look at examples which help us understand how the conditional model differs from the usual multivariate extreme value model.

Example 2.5.1. We start by considering the 2-dimensional non-negative orthant. This example emphasizes the fact that we need different normalizations for different cones. This is known for hidden regular variation with the cones $\mathbb{E}$ and $\mathbb{E}_0$ (Example 5.1 in Maulik and Resnick [2005]). We still need a different normalization for the cones $\mathbb{E}_{\sqcap}$ and $\mathbb{E}_{\sqsupset}$.

Let X and Z be i.i.d. $Pareto(1)$ random variables. Define $Y = X^2 \wedge Z^2$. Then it is easy to see that the following hold:

(i) In $\mathbb{M}_+(\mathbb{E})$

$$t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in ([0,x] \times [0,y])^c\Big] \to \frac{1}{x} + \frac{1}{y}, \quad x \vee y > 0. \tag{2.5.1}$$

(ii) In $\mathbb{M}_+(\mathbb{E}_0)$: For $\frac{1}{2} < \alpha < 1$,

$$t\mathbf{P}\Big[\Big(\frac{X}{t^\alpha}, \frac{Y}{t^{2(1-\alpha)}}\Big) \in (x,\infty] \times (y,\infty]\Big] \to \frac{1}{x\sqrt{y}}, \quad x \wedge y > 0,$$

or in standard form,

$$t\mathbf{P}\Big[\Big(\frac{X^{1/\alpha}}{t}, \frac{Y^{1/2(1-\alpha)}}{t}\Big) \in (x,\infty] \times (y,\infty]\Big] \to \frac{1}{x^\alpha y^{1-\alpha}}, \quad x \wedge y > 0, \tag{2.5.2}$$

(iii) In $\mathbb{M}_+(\mathbb{E}_{\sqcap})$, the limit is not a product measure,

$$t\mathbf{P}\Big[\Big(\frac{X}{t^{1/2}}, \frac{Y}{t}\Big) \in [0,x] \times (y,\infty]\Big] \to \frac{1}{y} - \frac{1}{\sqrt{y}} \times \frac{1}{x \vee \sqrt{y}}, \quad x \geq 0, y > 0,$$

so a standard form exists,

$$tP\Big[\Big(\frac{X^2}{t},\frac{Y}{t}\Big)\in[0,x]\times(y,\infty]\Big]\to\frac{1}{y}-\frac{1}{\sqrt{y}}\times\frac{1}{\sqrt{x}\vee\sqrt{y}},\quad x\geq 0, y>0. \tag{2.5.3}$$

(iv) In $\mathbb{M}_+(\mathbb{E}_{\sqsupset})$, again, the limit is not a product measure,

$$tP\Big[\Big(\frac{X}{t},\frac{Y}{t^2}\Big)\in(x,\infty]\times[0,y]\Big]\to\frac{1}{x}-\frac{1}{x\vee\sqrt{y}},\quad x>0, y\geq 0,$$

so a standard form exists,

$$tP\Big[\Big(\frac{X}{t},\frac{Y^{1/2}}{t}\Big)\in(x,\infty]\times[0,y]\Big]\to\frac{1}{x}-\frac{1}{x\vee y},\quad x>0, y\geq 0. \tag{2.5.4}$$

These results can also be viewed in terms of polar co-ordinates by using the transformation $(r,\theta):(x,y)\mapsto(x+y,\frac{x}{x+y})$ (Section 2.3.4). Note that the absolute value of the Jacobian of the inverse transformation here is $|J|=r$. Hence,

$$f_{R,\Theta}(r,\theta)=rf_{X,Y}(r\theta,r(1-\theta)).$$

Let us look at the different cones in cases (i)–(iii).

(i) The angular measure has a point mass at 0 and 1,

$$S(d\theta)=\delta_{\{0\}}(d\theta)+\delta_{\{1\}}(d\theta).$$

(ii) The limit measure in standard form is

$$\mu((x,\infty]\times(y,\infty])=\frac{1}{x^{\alpha}y^{1-\alpha}},\quad x\wedge y>0$$

for $\frac{1}{2}<\alpha<1$. Hence,

$$\mu'(x,y)=\frac{\partial}{\partial x}\frac{\partial}{\partial y}\mu((x,\infty]\times(y,\infty])=\frac{\alpha(1-\alpha)}{x^{\alpha+1}y^{2-\alpha}}.$$

Taking the polar coordinate transformation

$$\mu'_{R,\Theta}(r,\theta) = r\frac{\alpha(1-\alpha)}{(r\theta)^{\alpha+1}(r(1-\theta))^{2-\alpha}} = r^{-2}\frac{\alpha(1-\alpha)}{\theta^{\alpha+1}(1-\theta)^{2-\alpha}}.$$

The right side is a product, as expected. Thus the angular measure has density

$$S(d\theta) = \frac{\alpha(1-\alpha)}{\theta^{\alpha+1}(1-\theta)^{2-\alpha}}\mathbf{1}_{\{0<\theta<1\}}d\theta.$$

(iii) The limit measure in standard form is

$$\mu([0,x]\times(y,\infty]) = \frac{1}{y} - \frac{1}{\sqrt{y}}\times\frac{1}{\sqrt{x}\vee\sqrt{y}}, \quad x \geq 0, y > 0,$$

which is equivalent to

$$\mu((x,\infty]\times(y,\infty]) = \frac{1}{\sqrt{y}}\times\frac{1}{\sqrt{x}\vee\sqrt{y}}, \quad x \geq 0, y > 0,$$

Hence, for $x > y > 0$

$$\mu'(x,y) = \frac{\partial}{\partial x}\frac{\partial}{\partial y}\mu((x,\infty]\times(y,\infty]) = \frac{1}{4}\frac{1}{x^{3/2}y^{3/2}}.$$

Taking the polar coordinate transformation, we get for $\theta > 1/2$,

$$\mu'_{R,\Theta}(r,\theta) = r\frac{1}{4}\frac{1}{(r\theta)^{3/2}(r(1-\theta))^{3/2}} = \frac{1}{4}r^{-2}\frac{1}{\theta^{3/2}(1-\theta)^{3/2}},$$

the density of a product measure. For $x \leq y$ the density does not exist and we have a point mass at $\theta = \frac{1}{2}$ whose weight can be calculated using (2.3.20). Thus the angular measure has density,

$$S(d\theta) = (2-\sqrt{3})\delta_{\{1/2\}}(d\theta) + \frac{1}{4}\theta^{-3/2}(1-\theta)^{-3/2}\mathbf{1}_{\{1/2<\theta<1\}}d\theta$$

(iv) The angular measure has a point mass at $\frac{1}{2}$,

$$S(d\theta) = 2\delta_{\{1/2\}}(d\theta).$$

Example 2.5.2. Suppose in Definition 2.1.1 we have functions $\alpha(t) > 0, \beta(t) = 0$ such that

$$\lim_{t\to\infty} \frac{\alpha(tc)}{\alpha(t)} = \psi_1(c) = c^\rho, \qquad \lim_{t\to\infty} \frac{\beta(tc) - \beta(t)}{\alpha(t)} = \psi_2(c) = 0,$$

for some $\rho \neq 0$. Refer to [Heffernan and Resnick, 2007, Remark 2, page 545]. In such a case, the limit measure μ satisfies:

$$\mu([-\infty, x] \times (y, \infty]) = y^{-1}\mu([-\infty, \frac{x}{y^\rho}] \times (1, \infty]) = y^{-1}H\big(\frac{x}{y^\rho}\big) \tag{2.5.5}$$

for $x \in \mathbb{R}$ and $y > 0$, where $H(\cdot)$ is a proper non-degenerate distribution. The following is an example of such a limit measure.

Assume $0 < \rho < 1$ and suppose $X \sim Pareto(\rho)$ and $Z \sim Pareto(1-\rho)$ are independent random variables. Define $Y = X \wedge Z$ and we have,

$$\begin{aligned} t\mathbf{P}\Big[\Big(\frac{X}{t}, \frac{Y}{t}\Big) \in [0, x] \times (y, \infty]\Big] &= t\mathbf{P}\Big(\frac{X}{t} \le x, \frac{X}{t} > y, \frac{Z}{t} > y\Big) \\ &= \frac{1}{y^{1-\rho}}\Big(\frac{1}{y^\rho} - \frac{1}{x^\rho}\Big), (\text{ for } x \ge y > 0 \text{ and } t \text{ large}) \\ &= \frac{1}{y}\Big(1 - \frac{y^\rho}{x^\rho}\Big) =: \mu^{**}([-\infty, x] \times (y, \infty]). \end{aligned}$$

Now as in Proposition 2.3.3, case 1, we have

$$\begin{aligned} t\mathbf{P}\Big[\Big(\frac{\alpha(X)}{\alpha(t)}, \frac{Y}{t}\Big) \in [0, x] \times (y, \infty]\Big] &\to \mu^{**}([0, x^{1/\rho}] \times (y, \infty]) \\ &= \frac{1}{y}\Big(1 - \frac{y^\rho}{x}\Big), \quad x \ge y^\rho > 0 \\ &=: \mu([0, x] \times (y, \infty]). \end{aligned}$$

If we take $H(\cdot)$ to be $Pareto(1)$, then we have the limit measure for $x \ge 0, y > 0$,

$$\mu([-\infty, x] \times (y, \infty]) \ := \frac{1}{y}H\big(\frac{x}{y^\rho}\big).$$

Example 2.5.3. This example provides us with a class of limit distributions on $\mathbb{E}_\sqcap$ that can be indexed by distributions on $[0, \infty]$. Suppose R is a Pareto random variable on $[1,\infty)$ with parameter 1 and ξ is a random variable with distribution $G(\cdot)$ on $[0,\infty]$. Assume that ξ and R are independent. Define the bivariate random vector $(X,Y) \in \mathbb{R}_+^2$ as

$$(X,Y) = (R\xi, R).$$

Therefore we have for $y > 0, x \geq 0$ (and $ty > 1$),

$$\begin{aligned}
t\mathbf{P}\Big[\frac{X}{t} \leq x, \frac{Y}{t} > y\Big] &= t\mathbf{P}\Big[\frac{R\xi}{t} \leq x, \frac{R}{t} > y\Big] = t\int\limits_{ty}^{\infty} \mathbf{P}\big[\xi \leq \frac{tx}{r}\big] r^{-2} dr \\
&= \int\limits_{y}^{\infty} \mathbf{P}\big[\xi \leq \frac{x}{s}\big] s^{-2} ds = \int\limits_{y}^{\infty} G\big(\frac{x}{s}\big) s^{-2} ds \qquad (\text{ putting } s = \frac{r}{t}) \\
&= \frac{1}{x}\int\limits_{0}^{x/y} G(s) ds = \mu\big([0,x] \times (y,\infty]\big).
\end{aligned}$$

This can be viewed in terms of polar co-ordinates. We know that an angular measure $S(\cdot)$ on $\mathbb{E}_{\sqcap}$ for $0 \leq \eta < 1$ can be given by

$$S([0,\eta]) = \mu\{(u,v) : u+v > 1, \frac{u}{u+v} \leq \xi\}.$$

Hence we have

$$\begin{aligned}
t\mathbf{P}\Big[\frac{X+Y}{t} > 1, \frac{X}{X+Y} \leq \eta\Big] &= t\mathbf{P}\Big[\frac{R\xi + R}{t} > 1, \frac{R\xi}{R\xi + R} \leq \eta\Big] \\
&= t\mathbf{P}\Big[\frac{R(1+\xi)}{t} > 1, \xi \leq \frac{\eta}{1-\eta}\Big] \\
&= t \int\limits_{0 \leq s \leq \frac{\eta}{1-\eta}} \mathbf{P}\Big[\frac{R}{t}(1+s) > 1\Big] G(ds) \\
&= t \int\limits_{0 \leq s \leq \frac{\eta}{1-\eta}} \Big(\frac{t}{1+s} \vee 1\Big)^{-1} G(ds) \\
&= \int\limits_{0 \leq s \leq \frac{\eta}{1-\eta}} (1+s) G(ds). \qquad \text{for } t > \frac{1}{1-\eta}
\end{aligned}$$

But the left side in the previous equation goes to $\mu\{(u,v) : u+v > 1, \frac{y}{u+v} \leq \xi\} = S([0,\eta])$ as $t \to \infty$. Hence we have

$$S([0,\eta]) = \int_{0 \leq s \leq \frac{\eta}{1-\eta}} (1+s)G(ds), \qquad 0 \leq \eta < 1.$$

Hence S is a finite angular measure if and only if G has first moment.

CHAPTER 3

DETECTION OF A CONDITIONAL EXTREME VALUE MODEL

3.1 Introduction

In Chapter 2 we provided some insight into the conditional extreme value model, primarily with respect to model consistency issues, relationship with MEVT and connection to regular variation on cones of the Euclidean space. Subsequently, it would be nice to know situations where the use of CEV modeling would be appropriate. It is known that in the presence of asymptotic independence, the limit measure in the multivariate EVT set up has an empty interior [Resnick, 2007, Chapter 6]; in other words, the limit measure concentrates on the boundary of the state space. In such a case an additional assumption of a CEV model provides more insight into the dependence structure. The CEV model also provides a way for modeling multivariate data assuming a subset rather than the entire vector to be extreme-valued. In this chapter we suggest situations where a CEV model can be used and suggest techniques to detect the model statistically and in the process also detect properties of the limit measure for the model. The methodologies are suggested for a bivariate data set.

Section 3.1 provides an introduction and review of the model. Section 3.2 deals with the detection of a conditional extreme value model. It has been shown in Chapter 2 that the CEV model can be standardized to regular variation on a special cone if and only if the limit measure involved is not a product. In case of a product measure in the limit, we need to estimate fewer parameters and calculating probabilities is also simpler. Fougères and Soulier [2008] suggests some estimates for parameters and normalizing constants in the two different cases (product and non-product limit measures). Hence it is important to know whether we are in the product case or not. In Section 3.2 we propose three statistics whose behavior can first of all indicate the appropriateness of the CEV model and secondly indicate whether the limit measure is a product or not. Section 3.3 is dedicated to applying our techniques to some simulated and real data coming from Internet traffic studies.

3.1.1 The CEV model

Section 2.1.1 in Chapter 2 provides the necessary preliminary results on the CEV model. Refer to Heffernan and Resnick [2007] and Chapter 2 of this thesis for further discussion on conditional extreme value models.

Now note that the CEV model primarily differs from the multivariate extreme value model in the domain of attraction condition. In Chapter 2 we have seen conditions under which a CEV model can be extended to multivariate extreme value model. Under the multivariate extreme value model, each of the variables can be standardized so that we have a multivariate regular variation on the cone $[\mathbf{0}, \boldsymbol{\infty}] \setminus \{\mathbf{0}\}$; see de Haan and Resnick [1977] and Chapter 6 of de Haan and Ferreira [2006]. The conditional extreme value model can be standardized if and only if the limit measure μ in (2.1.3) is not a product measure (Proposition 2.3.3). When both X and Y are standardized, we can characterize the limit measure in terms of all Radon measures (finite and infinite) on $[0, 1)$. Though theoretically elegant, performing standardization in practice is not an easy task.

Thus, it is important to know when the limit is a product measure. A product measure in the limit precludes standardization of both the variables [Heffernan and Resnick, 2007] and means that we do not have a multivariate extreme value model (Proposition 2.3.3). However, a product measure makes the estimation of certain parameters and probabilities easier. For instance, in the product case $(\psi_1, \psi_2) \equiv (1, 0)$ so $\rho = 0$ (see (2.1.10)) but without the property that μ is a product, ρ has to be estimated. Furthermore, the limit being a product measure can be considered as a form of *asymptotic independence* in the CEV model, which can be probabilistically useful [Maulik et al., 2002].

3.1.2 Appropriateness of the CEV model.

Multivariate extreme value theory provides a rich literature on estimation of probabilities of extreme regions containing few or no data points in the sample. The multivariate theory assumes that each variable is marginally in an extreme value domain of attraction. However, this might not be the right assumption for all data sets. We encounter data where one or some but not all the variables can be assumed to be in an extreme value domain; see Section 3.3.2. The CEV model is a candidate model in such cases.

Another circumstance where the CEV model can be helpful is if one has a multivariate extreme value model with limit measure ν possessing *asymptotic independence*. This means that in the standardized model, the limit measure, $\nu^*(\cdot)$, concentrates on the axes through $\{\mathbf{0}\}$ and $\nu^*((\mathbf{0}, \boldsymbol{\infty}]) = 0$. So an estimate of the probability of a region where both variables are big will turn out to be zero which may be a useless and misleading estimate. In such a circumstance, finer estimates can be obtained using either *hidden regular variation* [Maulik and

Resnick, 2005] or the CEV model. Both methods provide a non-zero limit measure by using normalization functions which are of different order from the multivariate EV model.

So, how do we decide if the CEV model is appropriate for multivariate data?

1. Start by checking whether any of the marginal variables belongs to the domain of attraction of an extreme value distribution. An informal way to do this is through plots of the estimators of the extreme value parameter γ, that is, Pickands plot, Moment estimator plot, etc [Embrechts et al., 1997, Resnick, 2007]. If the plot attains stability in some range it is reasonable to assume an extreme-value model. More formal methods for testing membership in a domain of attraction using quantile and distribution functions are discussed in de Haan and Ferreira [2006], Chapter 5.2. The special case of a heavy-tailed random variable can be detected using the QQ plot, plotting the theoretical quantiles of the exponential distribution versus the logarithm of the sorted data and checking for linearity in the high values of the data. This is reviewed in Resnick [2007]. We will deal with more of this in Chapter 4.

2. If some, but not all, marginal variables are in a domain of attraction, proceed to see if the data is consistent with the CEV model. See Section 3.2.

3. If all variables are in some extreme value domain, check if the multivariate extreme value model is appropriate and if asymptotic independence is present. One way to do this is by checking whether both maximum and minimum of the standardized variables have distributions with regularly varying tails [Coles et al., 1999, Resnick, 2002]. If the EV model is appropriate and asymptotic independence is absent, the CEV model does not

provide any more information than the EV model. On the other hand if asymptotic independence is present, the CEV model, if detected, provides supplementary information about the joint behavior of the variables away from at least one of the axes.

3.2 Three estimators for detecting the CEV model

Let $(X_1, Y_1), \ldots, (X_n, Y_n)$ be a bivariate random sample. In this section we propose three statistics to detect whether our sample is consistent with the CEV model under the assumption that at least one of the variables is in an extreme-value domain, and without loss of generality we assume Y to be that variable. Our statistics have a consistency property which allows detection of a product form for the limit measure.

Assume $(X_1, Y_1), \ldots, (X_n, Y_n)$ is i.i.d. from a CEV model as defined in Section 2.1.1. We first formulate a consequence of (2.1.3) which will be convenient for our purpose. The following notations will be used.

$Y_{(1)} \geq \ldots \geq Y_{(n)}$	The decreasing order statistics of $Y_1, \ldots, Y_n$.
X_i^*, for $1 \leq i \leq n$	The X-variable corresponding to $Y_{(i)}$, also called the concomitant of $Y_{(i)}$.
$R_i^k = \sum_{l=i}^{k} \mathbf{1}_{\{X_l^* \leq X_i^*\}}$	Rank of X_i^* among $X_1^*, \ldots, X_k^*$. For convenience we write $R_i = R_i^k$.
$X_{1:k}^* \leq X_{2:k}^* \leq \ldots \leq X_{k:k}^*$	The increasing order statistics of $X_1^*, \ldots, X_k^*$.

3.2.1 A consequence for empirical measures

When the CEV property holds, a family of point processes of ranks of the sample converge vaguely to a Radon measure. By transforming to ranks of the data, we presumably lose efficiency since only the relative ordering in the sample remains unchanged but detection of the CEV property is easier since we no longer need to estimate the various parameters of the model. See de Haan and de Ronde [1998], de Haan and Ferreira [2006], Resnick [2007].

The convergence statement (2.1.3) of the CEV model defined in Section 2.1.1 can be interpreted in terms of vague convergence of measures. In preparation for the forthcoming result we recall some commonly used notation and concepts. Let $\mathbb{E}^*$ be a locally compact space with a countable base (for example, a finite dimensional Euclidean space). We denote by $\mathbb{M}_+(\mathbb{E}^*)$, the non-negative Radon measures on Borel subsets of $\mathbb{E}^*$. If $\mu_n \in \mathbb{M}_+(\mathbb{E}^*)$ for $n \geq 0$, then μ_n converges vaguely to μ_0 (written $\mu_n \xrightarrow{v} \mu_0$) if for all bounded continuous functions f with compact support we have

$$\int_{\mathbb{E}^*} f d\mu_n \to \int_{\mathbb{E}^*} f d\mu_0 \quad (n \to \infty).$$

This concept allows us to write (2.1.3) as

$$t\mathbf{P}\left(\Big(\frac{X - \beta(t)}{\alpha(t)}, \frac{Y - b(t)}{a(t)} \Big) \in \cdot \right) \xrightarrow{v} \mu(\cdot), \text{ as } t \to \infty \tag{3.2.1}$$

in $\mathbb{M}_+([-\infty, \infty] \times \overline{\mathbb{E}}^{(\gamma)})$. Standard references include Kallenberg [1983], Neveu [1977] and [Resnick, 2008b, Chapter 3].

Recall the definition of μ^* in (2.1.9) and define the measure $L(\cdot) \in \mathbb{M}_+([0,1] \times [1, \infty])$ by

$$L([0, x] \times (y, \infty]) = \mu^*([-\infty, H^{\leftarrow}(x)] \times [y, \infty]), \quad (x, y) \in [0, 1] \times (1, \infty]. \tag{3.2.2}$$

Applying a reciprocal transformation to the second coordinate,

$$T_0 : (x, y) \mapsto (x, y^{-1})$$

converts L into the copula $L \circ T_0^{-1}$.

Proposition 3.2.1. *Suppose* $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ *are i.i.d. observations from a CEV model which follows* (2.1.2)-(2.1.6) *and suppose* H *is continuous. If* $k = k(n) \to \infty, n \to \infty$ *with* $k/n \to 0$*, then in* $\mathbb{M}_+([0,1] \times [1,\infty])$

$$\frac{1}{k}\sum_{i=1}^{k} \epsilon_{\left(\frac{R_i}{k}, \frac{k+1}{i}\right)}(\cdot) \Rightarrow L(\cdot).$$

Proof. From (3.2.1) and [Resnick, 2007, Theorem 5.3(ii)], as $n, k \to \infty$ with $\frac{k}{n} \to 0$,

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{X_i - \beta(n/k)}{\alpha(n/k)}, \frac{Y_i - b(n/k)}{a(n/k)}\right)}(\cdot) \Rightarrow \mu(\cdot), \tag{3.2.3}$$

in $\mathbb{M}_+([-\infty,\infty] \times \overline{\mathbb{E}}^{(\gamma)})$. Recall $Y_{(1)} \geq Y_{(2)} \geq \ldots \geq Y_{(n)}$ are the order statistics of $Y_1, \ldots, Y_n$ in decreasing order and ordering the Y's in (3.2.3) allows us to write the equivalent statement

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}, \frac{Y_{(i)} - b(n/k)}{a(n/k)}\right)}(\cdot) \Rightarrow \mu(\cdot). \tag{3.2.4}$$

Define the measure ν_γ by

$$\nu_\gamma\big((y,\infty] \cap \overline{\mathbb{E}}^{(\gamma)}\big) = (1+\gamma y)^{-1/\gamma}, \quad y \in \mathbb{E}^{(\gamma)},$$

and sometimes, here and elsewhere, we sloppily write $\nu_\gamma(y,\infty]$. Taking marginal convergence in (3.2.3), or using (2.1.2), we have with that

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\frac{Y_i - b(n/k)}{a(n/k)}}(\cdot) \Rightarrow \nu_\gamma(\cdot),$$

in $\mathbb{M}_+(\overline{\mathbb{E}}^{(\gamma)}$. Using an inversion technique (Resnick and Stărică [1995], de Haan and Ferreira [2006], [Resnick, 2007, page 82]), we get

$$\frac{Y_{(\lceil (k+1)t \rceil)} - b(n/k)}{a(n/k)} \xrightarrow{P} \frac{t^{-\gamma} - 1}{\gamma}, \tag{3.2.5}$$

in $D_l\big((0,\infty],\overline{\mathbb{E}}^{(\gamma)}\big)$, the class of left continuous functions on $(0,\infty]$ with range $\overline{\mathbb{E}}^{(\gamma)}$ and with finite right limits on $(0,\infty)$. The convergence in (3.2.5) being to a non-random function, we can append it to the convergence in (3.2.4) to get the following [Billingsley, 1968, p.27]:

$$\begin{aligned}(\mu_n, x_n(t) :=&\left(\frac{1}{k}\sum_{i=1}^{n}\epsilon_{\left(\frac{X_i^*-\beta(n/k)}{\alpha(n/k)},\frac{Y_{(i)}-b(n/k)}{a(n/k)}\right)}(\cdot), \frac{Y_{(\lceil (k+1)t\rceil)}-b(n/k)}{a(n/k)}\right)\\ &\Rightarrow\Big(\mu(\cdot),\frac{t^{-\gamma}-1}{\gamma}\Big) = (\mu, x_\infty(t))\end{aligned} \tag{3.2.6}$$

in $\mathbb{M}_+([-\infty,\infty]\times\overline{\mathbb{E}}^{(\gamma)})\times D_l\big((0,\infty],\overline{\mathbb{E}}^{(\gamma)})\big)$.

Let $D_l^{\downarrow}\big((0,\infty],\overline{\mathbb{E}}^{(\gamma)})\big)$ be the subfamily of $D_l\big((0,\infty],\overline{\mathbb{E}}^{(\gamma)})\big)$ consisting of non-increasing functions and define

$$T_1:\mathbb{M}_+([-\infty,\infty]\times\overline{\mathbb{E}}^{(\gamma)})\times D_l^{\downarrow}\big((0,\infty],\overline{\mathbb{E}}^{(\gamma)})\big)\mapsto\mathbb{M}_+\big([-\infty,\infty]\times(0,\infty]\big)$$

by

$$T_1(m, x(\cdot)) = m^*$$

where

$$m^*([-\infty,x]\times(t,\infty]) = m([-\infty,x]\times(x(t^{-1}),\infty]),\quad x\in[-\infty,\infty],\, t\in(0,\infty].$$

This is an a.s. continuous map so apply this to (3.2.6) and

$$T_1(\mu_n, x_n)\Rightarrow T_1(\mu, x_\infty). \tag{3.2.7}$$

For $x\in[-\infty,\infty]$ and $y\in(0,\infty]$, the left side of (3.2.7) on the set $[-\infty,x]\times(y,\infty]$ is

$$\mu_n\big([-\infty,x]\times(\frac{Y_{(\lceil (k+1)t\rceil)}-b(n/k)}{a(n/k)},\infty]\big)$$

and since

$$\frac{Y_{(i)}-b(n/k)}{a(n/k)} > \frac{Y_{(\lceil (k+1)t\rceil)}-b(n/k)}{a(n/k)}$$

iff

$$i < (k+1)y^{-1} \quad \text{or} \quad \frac{k+1}{i} > y,$$

the left side of (3.2.7) on the set $[-\infty, x] \times (y, \infty]$ is

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}, \frac{k+1}{i}\right)}([-\infty, x] \times (y, \infty]).$$

The right side of (3.2.7) on the set $[-\infty, x] \times (y, \infty]$ is

$$\mu\big([-\infty, x] \times (\frac{y^\gamma - 1}{\gamma}, \infty]\big) = \mu^*\big([-\infty, x] \times (y, \infty]\big),$$

so we conclude

$$\frac{1}{k}\sum_{i=1}^{n} \epsilon_{\left(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}, \frac{k+1}{i}\right)} \Rightarrow \mu^* \tag{3.2.8}$$

in $\mathbb{M}_+\big([-\infty, \infty] \times (0, \infty]\big)$. Recall $\mu^*(\cdot)$ was defined in (2.1.9).

Now assuming $(x, 1)$ is a continuity point of μ^* we have

$$\begin{aligned} H_n(x) :=& \frac{1}{k}\sum_{i=1}^{k} \epsilon_{\left(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}, \frac{k+1}{i}\right)}([-\infty, x] \times (1, \infty]) \\ \Rightarrow& \mu^*([-\infty, x] \times (1, \infty]) =: H(x), \end{aligned} \tag{3.2.9}$$

or in the topology of weak convergence on $PM[-\infty, \infty]$, the probability measures on $[-\infty, \infty]$,

$$H_n \Rightarrow H.$$

Define a map T_2 on $\mathbb{M}_+([-\infty, \infty] \times [1, \infty]) \times PM[-\infty, \infty]$ by

$$T_2(m, G) = m^\#$$

where

$$m^\#([0, z] \times (y, \infty]) = m([0, G(z)] \times (y, \infty])$$

or, for $f \in C([0, 1] \times [1, \infty])$,

$$m^\#(f) = \iint f(G(x), y) m(dx, dy).$$

This map is continuous at (m, G) provided G is continuous. To see this, let f be continuous on $[0,1] \times [1,\infty]$ and suppose $G_n \Rightarrow G$ and $m_n \xrightarrow{v} m$. Then

$$\begin{aligned}
\Big|\iint f(G_n(x), y) m_n(dx, dy) - \iint f(G(x), y) m_n(dx, dy)\Big| \\
\leq & \Big|\iint f(G_n(x), y) m_n(dx, dy) - \iint f(G(x), y) m_n(dx, dy)\Big| \\
& \quad + \Big|\iint f(G(x), y) m_n(dx, dy) - \iint f(G(x), y) m(dx, dy)\Big| \\
= & I + II.
\end{aligned}$$

For I, convergence to 0 follows by uniform continuity of f and the fact that $G_n(x) \to G(x)$ uniformly in x. To verify $II \to 0$, it suffices to note that $f(G(x), y)$ is continuous with compact support $[-\infty, \infty] \times [1, \infty]$ and then use $m_n \xrightarrow{v} m$.

Combine (3.2.8) and (3.2.9) to get

$$\Big(\frac{1}{k}\sum_{i=1}^{k} \epsilon_{\big(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}, \frac{k+1}{i}\big)}, H_n\Big) \Rightarrow (\mu^*, H) \tag{3.2.10}$$

in $\mathbb{M}_+\big([-\infty,\infty] \times [1,\infty]\big) \times PM[-\infty,\infty]$. Apply the transformation T_2 discussed in the previous paragraph. The limit at $[0,x] \times (y,\infty]$ is $\mu^*\big([-\infty, H^{\leftarrow}(x)] \times (y,\infty]\big) = L[0,x] \times (y,\infty]$. The converging sequence can be written as

$$\frac{1}{k}\sum_{i=1}^{k} \epsilon_{\Big(H_n\big(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}\big), \frac{k+1}{i}\Big)}.$$

Finally observe

$$H_n\big(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}\big) = \frac{1}{k}\sum_{l=1}^{k} 1_{[\frac{X_l^* - \beta(n/k)}{\alpha(n/k)} \leq \frac{X_i^* - \beta(n/k)}{\alpha(n/k)}]} = \frac{R_i}{k}.$$

The result follows. □

We propose three statistics that can be used to detect whether or not a CEV model is appropriate, and if so, whether the model has a product measure in the limit.

3.2.2 The Hillish statistic, Hillish$_{k,n}$

The Hill estimator (Hill [1975], Mason [1982], de Haan and Ferreira [2006], Resnick [2007]) is a popular choice for estimating the tail parameter α of a heavy-tailed distribution. We say that a distribution function F on $\mathbb{R}$ is heavy-tailed with tail parameter $\alpha > 0$ if

$$\lim_{t\to\infty} \frac{1-F(tx)}{1-F(t)} = x^{-\alpha} \qquad \text{for} \quad x > 0. \tag{3.2.11}$$

If $Z_1, Z_2, \ldots, Z_n$ are i.i.d from this distribution F and $Z_{(1)} \geq Z_{(2)} \geq \ldots \geq Z_{(n)}$ are the orders statistics of the sample in decreasing order, then the Hill estimator defined as

$$\text{Hill}_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log \frac{Z_{(i)}}{Z_{(k+1)}}$$

is a weakly consistent estimator of $\frac{1}{\alpha}$ as $k, n \to \infty, k/n \to 0$. One way to obtain the consistency is to integrate the tail empirical measure and use its consistency. See Resnick and Stărică [1995] or [Resnick, 2007, p. 81].

The Hillish statistic, based on the ranks of the sample, converges weakly to a constant limit under the CEV model. The name is derived from the similarity of proof of this convergence with that of the weak consistency of the Hill estimator. Using the notation defined just prior to Section 2.1, and assuming $(\boldsymbol{X}, \boldsymbol{Y}) := \{(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)\}$, the Hillish statistic for $(\boldsymbol{X}, \boldsymbol{Y})$ is defined as

$$\text{Hillish}_{k,n}(\boldsymbol{X}, \boldsymbol{Y}) := \frac{1}{k}\sum_{j=1}^{k} \log\frac{k}{R_j} \log\frac{k}{j}. \tag{3.2.12}$$

The following proposition provides a convergence result of the Hillish statistic under conditions on k.

Proposition 3.2.2. *Suppose* $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ *are i.i.d. observations from a CEV model which follows* (2.1.2)-(2.1.6) *and suppose* H *as defined in* (2.1.6) *is continuous. Assume that* $k = k(n) \to \infty$, $n \to \infty$ *and* $k/n \to 0$. *Then*

$$\text{Hillish}_{k,n} \xrightarrow{P} \int_1^\infty \int_1^\infty \mu^*([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y, \infty)) \frac{dx}{x} \frac{dy}{y} =: I_{\mu^*}. \tag{3.2.13}$$

Proof. Proposition 3.2.1 yields

$$\frac{1}{k} \sum_{i=1}^{n} \epsilon_{\left(\frac{R_i}{k}, \frac{k+1}{i}\right)}(\cdot) \Rightarrow L(\cdot) \tag{3.2.14}$$

in $\mathbb{M}_+\big([0,1] \times [1,\infty]\big)$. Rewrite (3.2.14) for $x \geq 1, y > 1$ as

$$\begin{aligned}\mu_n^*([x,\infty] \times (y,\infty)) :=& \frac{1}{k} \sum_{i=1}^{n} \epsilon_{\left(\frac{k}{R_i}, \frac{k+1}{i}\right)}([x,\infty] \times (y,\infty)) \\ &\Rightarrow \mu^*([-\infty, H^{\leftarrow}(1/x)] \times (y,\infty]).\end{aligned} \tag{3.2.15}$$

Observe that

$$\begin{aligned}I_n :=& \int_1^\infty \int_1^\infty \mu_n^*([x,\infty] \times (y,\infty)) \frac{dx}{x} \frac{dy}{y} \\ =& \frac{1}{k} \int_1^\infty \int_1^\infty \sum_{i=1}^{n} \epsilon_{\left(\frac{k}{R_i}, \frac{k+1}{i}\right)}([x,\infty] \times (y,\infty)) \frac{dx}{x} \frac{dy}{y} \\ =& \frac{1}{k} \sum_{i=1}^{k} \log \frac{k}{R_i} \log \frac{k+1}{i} = \frac{1}{k} \sum_{i=1}^{k} \log \frac{k}{R_i} (\log \frac{k}{i} + \log \frac{k+1}{k}) \\ =& \frac{1}{k} \sum_{i=1}^{k} \log \frac{k}{R_i} \log \frac{k}{i} + \big(\log \frac{k+1}{k}\big) \frac{1}{k} \sum_{i=1}^{k} \log \frac{k}{R_i} \\ =& \text{Hillish}_{k,n} + A_k\end{aligned} \tag{3.2.16}$$

where $A_k := \big(\log \frac{k}{k+1}\big) \frac{1}{k} \sum_{i=1}^{k} \log \frac{k}{i} \to 0 \times 1 = 0$ as $k \to \infty$. Hence if we show

$$I_n \xrightarrow{P} \int_1^\infty \int_1^\infty \mu^*([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y,\infty)) \frac{dx}{x} \frac{dy}{y},$$

then we are done. For N finite we know that

$$\int_1^N \int_1^N \mu_n^*([x,\infty] \times (y,\infty]) \frac{dx}{x}\frac{dy}{y}$$

$$\xrightarrow{P} \int_1^N \int_1^N \mu^*([-\infty, H^{\leftarrow}(1/x)] \times (y,\infty]) \frac{dx}{x}\frac{dy}{y}, \tag{3.2.17}$$

since (3.2.15) implies that the integrand converges in probability and we can use Pratt's Lemma [Resnick, 1999, page 164] for the convergence of the integral. Note that as $N \to \infty$ the right hand side in equation (3.2.17) converges to $\int_1^\infty \int_1^\infty \mu^*([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y,\infty]) \frac{dx}{x}\frac{dy}{y}$. So we need to see what happens outside the compact sets. In particular if we can show that for any $\delta > 0$,

$$\lim_{N\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(\int_N^\infty \int_1^\infty \mu_n^*([x,\infty] \times (y,\infty]) \frac{dx}{x}\frac{dy}{y} > \delta \right) = 0 \tag{3.2.18}$$

and

$$\lim_{N\to\infty} \limsup_{n\to\infty} \mathbf{P}\left(\int_1^\infty \int_N^\infty \mu_n^*([x,\infty] \times (y,\infty]) \frac{dx}{x}\frac{dy}{y} > \delta \right) = 0 \tag{3.2.19}$$

then by a standard converging together theorem [Resnick, 2007, Theorem 3.5], we are done. Observe that

$$\begin{aligned} 0 \le \int_1^\infty \int_N^\infty \mu_n^*([x,\infty] \times (y,\infty]) \frac{dx}{x}\frac{dy}{y} &= \frac{1}{k}\sum_{j=1}^{k} \log\frac{k}{R_j}(\log 1 \vee \log\frac{k}{jN}) \\ &\le \sqrt{\frac{1}{k}\sum_{j=1}^{k}\big(\log\frac{k}{R_j}\big)^2 \frac{1}{k}\sum_{j=1}^{k}\big(\log 1 \vee \log\frac{k}{jN}\big)^2} \qquad \text{(Cauchy-Schwarz)} \\ &= B_{k_n} \times C_{k_n,N} \end{aligned}$$

where

$$B_{k_n}^2 = \frac{1}{k}\sum_{j=1}^{k}\big(\log\frac{k}{R_j}\big)^2 = \frac{1}{k}\sum_{j=1}^{k}\big(\log\frac{k}{j}\big)^2 \sim \int_0^1 (-\log x)^2 dx = 2$$

and

$$C^2_{k_n,N} = \frac{1}{k}\sum_{j=1}^{k}\big(0 \vee \log\frac{k}{jN}\big)^2 = \frac{1}{k}\sum_{j\leq k/N}\big(\log\frac{k}{jN}\big)^2$$
$$= \frac{1}{N}\frac{1}{k/N}\sum_{j=1}^{k/N}\big(\log\frac{k}{jN}\big)^2 \sim \frac{1}{N}\int\limits_0^1(-\log x)^2dx = \frac{1}{N}\times 2.$$

Hence

$$\limsup_{n\to\infty}\mathbf{P}\Big[\int\limits_1^\infty\int\limits_N^\infty \mu^*_n([x,\infty]\times(y,\infty])\frac{dx}{x}\frac{dy}{y} > \delta\Big] \leq \limsup_{n\to\infty}\mathbf{P}(B_{k_n}\times C_{k_n,N} > \delta)$$

and applying Fatou's Lemma, this is bounded by

$$\leq \mathbf{P}\Big[\sqrt{2\times\frac{2}{N}} > \delta\Big] \to 0 \qquad (N\to\infty).$$

This shows (3.2.19) holds and similarly we can show (3.2.18) holds, and we are done. $\square$

Suppose $(\boldsymbol{X},\boldsymbol{Y}) := \{(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)\}$ is a sample from a CEV limit model with normalizing functions α,β,a,b and variational functions ψ_1,ψ_2. Let the standardized limit measure be μ^* as defined in (2.1.9). Also $H(x) = \mu^*([-\infty,x]\times(1,\infty])$. Then $(-\boldsymbol{X},\boldsymbol{Y})$ is also a sample from a CEV limit model but with normalizing functions $\tilde{\alpha} = \alpha, \tilde{\beta} = -\beta, \tilde{a} = a, \tilde{b} = b$ and variational functions $\tilde{\psi}_1 = \psi_1, \tilde{\psi}_2 = -\psi_2$. In this case the standardized limit measure is $\tilde{\mu^*}$ and it is easy to check that for $x\in\mathbb{R}$, $y>0$,

$$\tilde{\mu^*}([-\infty,x]\times(y,\infty]) = \mu^*([-x,\infty]\times(y,\infty]). \tag{3.2.20}$$

We also have for $x\in\mathbb{R}$

$$\tilde{H}(x) := \tilde{\mu^*}([-\infty,x]\times(1,\infty]) = \mu^*([-x,\infty]\times(1,\infty]) = 1 - H(-x). \tag{3.2.21}$$

Thus, for $0 < p < 1$, we have,

$$\tilde{H}^{\leftarrow}(p) = -H^{\leftarrow}(1-p). \tag{3.2.22}$$

The following proposition characterizes product measure in terms of limits of the Hillish statistic for both $(\boldsymbol{X}, \boldsymbol{Y})$ and $(-\boldsymbol{X}, \boldsymbol{Y})$.

Proposition 3.2.3. *Under the conditions of Proposition 3.2.2, μ^* is a product measure if and only if both*

$$\mathit{Hillish}_{k,n}(\boldsymbol{X}, \boldsymbol{Y}) \xrightarrow{P} 1 \quad \mathit{and} \quad \mathit{Hillish}_{k,n}(-\boldsymbol{X}, \boldsymbol{Y}) \xrightarrow{P} 1.$$

Proof. Evaluating I_{μ^*}, the limit of the Hillish statistic as proposed in Proposition 3.2.2, leads us to the above results. Recall that we assume H is continuous. Define for any $c > 0$, the family $H^{(c)}(\cdot)$ of distribution functions as follows:

$$H^{(c)}(x) := c^{-1}\mu^*([-\infty, x] \times (c^{-1}, \infty]) = H(\psi_1(c)x + \psi_2(c))$$

where ψ_1, ψ_2 are as defined in (2.1.10) and the second equality can be obtained by using tc instead of t in the CEV model property (2.1.3) [Heffernan and Resnick, 2007, page 543]. Note that $H^{(1)} \equiv H$ according to our definition. Now,

$$\begin{aligned}\mu^*([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y, \infty]) =& \frac{1}{y} \times y\mu^*([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y, \infty]) \\ =& \frac{1}{y} \times H(\psi_1(1/y)H^{\leftarrow}(\frac{1}{x}) + \psi_2(1/y)).\end{aligned} \tag{3.2.23}$$

1. If μ^* is a product measure then $\mu^* = H \times \nu_1$ where $\nu_1\big((x, \infty]\big) = x^{-1}, x > 0$. Similarly, $\tilde{\mu^*} = \tilde{H} \times \nu_1$. We know that μ^* being a product measure is

equivalent to $\psi_1 \equiv 1, \psi_2 \equiv 0$. Thus $H^{(c)} \equiv H$ for any $c > 0$. Thus

$$
\begin{aligned}
I_{\mu^*} &= \int_1^\infty \int_1^\infty \mu^*\big([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y, \infty]\big) \frac{dx}{x}\frac{dy}{y} \\
&= \int_1^\infty \int_1^\infty \frac{1}{y} H\big(H^{\leftarrow}(\frac{1}{x})\big) \frac{dx}{x}\frac{dy}{y} \\
&= \int_1^\infty \int_1^\infty \frac{1}{y}\frac{1}{x}\frac{dx}{x}\frac{dy}{y} = \Big(\int_1^\infty \frac{1}{x^2} dx\Big)^2 = 1.
\end{aligned}
$$

Also

$$
\begin{aligned}
I_{\tilde{\mu^*}} &= \int_1^\infty \int_1^\infty \tilde{\mu^*}\big([-\infty, \tilde{H}^{\leftarrow}(\frac{1}{x})] \times (y, \infty]\big) \frac{dx}{x}\frac{dy}{y} \\
&= \int_1^\infty \int_1^\infty \mu^*\big([-\tilde{H}^{\leftarrow}(1\frac{1}{x})] \times (y, \infty]\big) \frac{dx}{x}\frac{dy}{y} \\
&= \int_1^\infty \int_1^\infty \frac{1}{y}\big(1 - H\big(H^{\leftarrow}(1 - \frac{1}{x})\big)\frac{dx}{x}\frac{dy}{y} \\
&= \Big(\int_1^\infty \frac{1}{x^2} dx\Big)^2 = 1.
\end{aligned}
$$

2. Conversely assume that $I_{\mu^*} = I_{\tilde{\mu^*}} = 1$. We know that $\psi_1(c) = c^\rho$ for some $\rho \in \mathbb{R}$. Let us consider the following cases:

   (a) $\rho = 0$. This means $\psi_1 \equiv 1$ and $\psi_2(c) = k \log c$ for some $k \in \mathbb{R}$. We will show that k must be 0. If $k > 0$, then

$$
\begin{aligned}
I_{\mu^*} &= \int_1^\infty \int_1^\infty \mu^*\big([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y, \infty]\big) \frac{dx}{x}\frac{dy}{y} \\
&= \int_1^\infty \int_1^\infty \frac{1}{y} H\big(H^{\leftarrow}(\frac{1}{x}) - k\log y)\big) \frac{dx}{x}\frac{dy}{y} \\
&< \int_1^\infty \int_1^\infty \frac{1}{y}\frac{1}{x}\frac{dx}{x}\frac{dy}{y} = \Big(\int_1^\infty \frac{1}{x^2} dx\Big)^2 = 1.
\end{aligned}
$$

Similarly, we can show

$$I_{\mu^*} \begin{cases} = 1 & \text{if} \quad k = 0 \\ > 1 & \text{if} \quad k < 0. \end{cases} \qquad \text{and} \qquad I_{\tilde{\mu^*}} \begin{cases} > 1 & \text{if} \quad k > 0 \\ = 1 & \text{if} \quad k = 0 \\ < 1 & \text{if} \quad k < 0. \end{cases}$$

Thus for $I_{\mu^*} = I_{\tilde{\mu^*}} = 1$ to hold, we must have $k = 0$, which implies $\psi_2 \equiv 0$ and μ^* becomes a product measure.

(b) $\rho \neq 0$. We will show that this is not possible under the assumption $I_{\mu^*} = I_{\tilde{\mu^*}} = 1$. For $c > 0$,

$$\psi_1(c) = c^\rho, \qquad \psi_2(c) = \frac{k}{\rho}(c^\rho - 1)$$

for some $k \in \mathbb{R}$. Assume first $\rho > 0$. Then $(\frac{1}{y})^\rho \leq 1$ for $y \geq 1$. Therefore, for such y,

$$(\frac{1}{y})^\rho H^{\leftarrow}(\frac{1}{x}) + \frac{k}{\rho}((\frac{1}{y})^\rho - 1) \leq H^{\leftarrow}(\frac{1}{x}) \quad \text{iff} \quad H^{\leftarrow}(\frac{1}{x}) + \frac{k}{\rho} \geq 0$$
$$\text{iff} \quad x \leq 1/H(-\frac{k}{\rho}) =: \delta, \qquad \delta \geq 1. \tag{3.2.24}$$

Denote

$$\chi(x,y) := H\big((\frac{1}{y})^\rho H^{\leftarrow}(\frac{1}{x}) + \frac{k}{\rho}((\frac{1}{y})^\rho - 1)\big), \qquad x \geq 1, y \geq 1. \tag{3.2.25}$$

Since H is non-decreasing

$$\begin{aligned}\frac{1}{y}\chi(x,y) &= \mu^*([-\infty, H^{\leftarrow}(\frac{1}{x})] \times (y, \infty]) \\ &= \frac{1}{y}H\big((\frac{1}{y})^\rho H^{\leftarrow}(\frac{1}{x}) + \frac{k}{\rho}((\frac{1}{y})^\rho - 1)\big) \\ &\leq \frac{1}{y}H\big(H^{\leftarrow}(\frac{1}{x})\big) = \frac{1}{x}\cdot\frac{1}{y} \qquad \text{iff } x \leq \delta, y \geq 1.\end{aligned} \tag{3.2.26}$$

Since $I_{\mu^*} = 1$, we have

$$\int_1^\infty \int_1^\infty \frac{1}{y}\chi(x,y)\frac{dx}{x}\frac{dy}{y} = 1 = \int_1^\infty \int_1^\infty \frac{1}{x}\frac{1}{y}\frac{dx}{x}\frac{dy}{y}. \tag{3.2.27}$$

We claim $1 < \delta < \infty$, since if δ is either 1 or ∞, then (3.2.26) and (3.2.27) imply that $\chi(x,y) = \frac{1}{x}$ almost everywhere which means

$$(\frac{1}{y})^{\rho} H^{\leftarrow}(\frac{1}{x}) + \frac{k}{\rho}((\frac{1}{y})^{\rho} - 1) = H^{\leftarrow}(\frac{1}{x})$$

which is impossible for all $y \geq 1$ when $\rho > 0$.

From (3.2.27) we have

$$\int_1^{\infty}\int_1^{\infty} \frac{1}{y}\Big[\frac{1}{x} - \chi(x,y)\Big]\frac{dx}{x}\frac{dy}{y} = 0,$$

that is,

$$\begin{aligned} &\int_1^{\infty}\int_1^{\delta} \frac{1}{y}\Big[\frac{1}{x} - \chi(x,y)\Big]\frac{dx}{x}\frac{dy}{y} \\ =&\int_1^{\infty}\int_{\delta}^{\infty} \frac{1}{y}\Big[\chi(x,y) - \frac{1}{x}\Big]\frac{dx}{x}\frac{dy}{y} = \Delta \text{ (say)}, \end{aligned} \tag{3.2.28}$$

where the integrands are non-negative on both sides using (3.2.26).

Now $I_{\tilde{\mu}^*} = 1$ implies that

$$\begin{aligned} 1 &= \int_1^{\infty}\int_1^{\infty} \tilde{\mu}^*([-\infty, \tilde{H}^{\leftarrow}(\frac{1}{x})] \times (y,\infty])\frac{dx}{x}\frac{dy}{y} \\ &= \int_1^{\infty}\int_1^{\infty} \tilde{\mu}^*([H^{\leftarrow}(1-\frac{1}{x}),\infty] \times (y,\infty])\frac{dx}{x}\frac{dy}{y} \\ &= \int_1^{\infty}\int_1^{\infty} \frac{1}{y}\Big(1 - \mu^*([-\infty, H^{\leftarrow}(1-\frac{1}{x})] \times (1,\infty])\Big)\frac{dx}{x}\frac{dy}{y} \\ &= \int_1^{\infty}\int_1^{\infty} \frac{1}{y}\Big(1 - H\Big(\psi_1(1/y)H^{\leftarrow}(1-\frac{1}{x}) + \psi_2(1/y)\Big)\Big)\frac{dx}{x}\frac{dy}{y} \\ &= \int_1^{\infty}\int_1^{\infty} \frac{1}{y}\Big[1 - \chi(\frac{x}{x-1}, y)\Big]\frac{dx}{x}\frac{dy}{y}. \end{aligned}$$

Use the transformation $z = \frac{x}{x-1}$ and the above equation becomes

$$
\begin{aligned}
1 &= \int_1^\infty \int_1^\infty \frac{1}{y}\frac{1}{z-1}\Big[1 - \chi(z,y)\Big]\frac{dz}{z}\frac{dy}{y} \\
&= \int_1^\infty \int_1^\infty \frac{1}{y}\frac{1}{z-1}\Big[\frac{1}{z} - \chi(z,y)\Big]\frac{dz}{z}\frac{dy}{y} + \int_1^\infty \int_1^\infty \frac{1}{y}\frac{1}{z-1}\Big[1 - \frac{1}{z}\Big]\frac{dz}{z}\frac{dy}{y} \\
&= \int_1^\infty \int_1^\infty \frac{1}{z-1}\frac{1}{y}\Big[\frac{1}{z} - \chi(z,y)\Big]\frac{dz}{z}\frac{dy}{y} + 1.
\end{aligned}
$$

Therefore we have

$$
\int_1^\infty \int_1^\infty \frac{1}{x-1}\frac{1}{y}\Big[\frac{1}{x} - \chi(x,y)\Big]\frac{dx}{x}\frac{dy}{y} = 0.
$$

Since $\chi(x,y) \le \frac{1}{x}$ if and only if $x \le \delta$ from (3.2.26), we have

$$
\begin{aligned}
&\int_1^\infty \int_1^\delta \frac{1}{x-1}\frac{1}{y}\Big[\frac{1}{x} - \chi(x,y)\Big]\frac{dx}{x}\frac{dy}{y} \\
&= \int_1^\infty \int_\delta^\infty \frac{1}{x-1}\frac{1}{y}\Big[\chi(x,y) - \frac{1}{x}\Big]\frac{dx}{x}\frac{dy}{y}
\end{aligned} \tag{3.2.29}
$$

where the integrands on both sides are non-negative. But referring to (3.2.28) we have

$$
\begin{aligned}
&\int_1^\infty \int_1^\delta \frac{1}{x-1}\frac{1}{y}\Big[\frac{1}{x} - \chi(x,y)\Big]\frac{dx}{x}\frac{dy}{y} \\
&\ge \frac{1}{\delta-1}\int_1^\infty \int_1^\delta \frac{1}{y}\Big[\frac{1}{x} - \chi(x,y)\Big]\frac{dx}{x}\frac{dy}{y} = \frac{\Delta}{\delta-1}
\end{aligned} \tag{3.2.30}
$$

with equality holding only if the integrand is 0 almost everywhere. Similarly we have

$$
\begin{aligned}
&\int_1^\infty \int_\delta^\infty \frac{1}{x-1}\frac{1}{y}\Big[\chi(x,y) - \frac{1}{x}\Big]\frac{dx}{x}\frac{dy}{y} \\
&\le \frac{1}{\delta-1}\int_1^\infty \int_\delta^\infty \frac{1}{y}\Big[\chi(x,y) - \frac{1}{x}\Big]\frac{dx}{x}\frac{dy}{y} = \frac{\Delta}{\delta-1}
\end{aligned} \tag{3.2.31}
$$

with equality holding only if the integrand is 0 almost everywhere. The integrand cannot be 0 since it will imply $\chi(x, y) = \frac{1}{x}$ almost everywhere meaning $\rho = 0$. But our assumption is $\rho > 0$. Thus with strict inequality holding for both (3.2.30) and (3.2.31) we have a contradiction in equation (3.2.29). Thus we cannot have $\rho > 0$.

The case with $\rho < 0$ can be proved similarly.

Hence the result. □

This corollary provides a detection technique for the limit measure being a product measure. Given a sample of size n, we plot $\text{Hillish}_{k,n}$ for values of k and then try to see whether it stabilizes close to 1 or not. If the statistic is close to another value, this is evidence that the model is applicable but the limit measure is not product.

3.2.3 The Pickandsish statistic, $\text{Pickandsish}_{k,n}(p)$

Another way to check the suitability of the CEV assumption and to detect a product measure in the limit is to use the Pickandsish statistic which is based on ratios of differences of ordered concomitants. The statistic is patterned on the Pickands estimate for the parameter of an extreme value distribution (Pickands [1975], [de Haan and Ferreira, 2006, page 83], [Resnick, 2007, page 93]). For a fixed $k < n$, recall that $X^*_{1:k} \leq \ldots \leq X^*_{k:k}$ are the order statistics in increasing order from $X^*_1, X^*_2, \ldots, X^*_k$, the concomitants of $Y_{(1)} \geq \ldots \geq Y_{(k)}$, the order statistics in decreasing order from $Y_1, Y_2, \ldots, Y_n$. For notational convenience for

$s \leq t$ write $X_{s:t} := X_{\lceil s \rceil : \lceil t \rceil}$. Now define the Pickandsish statistic for $0 < p < 1$,

$$\text{Pickandsish}_{k,n}(p) := \frac{X^*_{pk:k} - X^*_{pk/2:k/2}}{X^*_{pk:k} - X^*_{pk/2:k}}. \tag{3.2.32}$$

Proposition 3.2.4. *Suppose* $(X_1, Y_1), \ldots, (X_n, Y_n)$ *follows a CEV model. Let* $0 < p < 1$*, Then, as* $k, n \to \infty$ *with* $k/n \to 0$*, we have*

$$\textit{Pickandsish}_{k,n}(p) \xrightarrow{P} \frac{H^{\leftarrow}(p)(1 - 2^{\rho}) - \psi_2(2)}{H^{\leftarrow}(p) - H^{\leftarrow}(p/2)}, \tag{3.2.33}$$

provided $H^{\leftarrow}(p) - H^{\leftarrow}(p/2) \neq 0$*. Here* ψ_1 *and* ψ_2 *are defined in* (2.1.10) *and* $\rho = \log(\psi_1(c))/\log c$.

Proof. Since H_n in (3.2.9) is a probability distribution converging to the limit H, we may invert the convergence and obtain [Resnick, 2007, Proposition 2.2, page 20],

$$H_n^{\leftarrow}(z) \xrightarrow{P} H^{\leftarrow}(z)$$

for $0 < z < 1$ for which $H^{\leftarrow}$ is continuous. The convergence of $H_n^{\leftarrow}(\cdot)$ translates to

$$\begin{aligned} H_n^{\leftarrow}(z) &= \inf\{u \in \mathbb{R} : H_n(u) \geq z\} \\ &= \inf\{u \in \mathbb{R} : \sum_{i=1}^{k} \epsilon_{\left(\frac{X_i^* - \beta(n/k)}{\alpha(n/k)}\right)}[-\infty, u] \geq kz\} \\ &= \frac{X^*_{\lceil kz \rceil:k} - \beta(n/k)}{\alpha(n/k)} \Rightarrow H^{\leftarrow}(z) \end{aligned} \tag{3.2.34}$$

where $X^*_{1:k} \leq \ldots X^*_{k:k}$ are the increasing order statistics of the concomitants $X^*_1, \ldots X^*_k$.

From (3.2.34), we have, for $0 < p \leq 1$, if $k, n \to \infty$ and $k/n \to 0$,

$$\frac{X^*_{pk:k} - \beta(n/k)}{\alpha(n/k)} \xrightarrow{P} H^{\leftarrow}(p), \tag{3.2.35}$$

$$\frac{X^*_{(p/2)k:k} - \beta(n/k)}{\alpha(n/k)} \xrightarrow{P} H^{\leftarrow}(p/2), \tag{3.2.36}$$

$$\frac{X^*_{pk/2:k/2} - \beta(2n/k)}{\alpha(2n/k)} \xrightarrow{P} H^{\leftarrow}(p). \tag{3.2.37}$$

Also recall from (2.1.10) that

$$\lim_{t\to\infty} \frac{\alpha(tc)}{\alpha(t)} = \psi_1(c) = c^\rho, \qquad \lim_{t\to\infty} \frac{\beta(tc) - \beta(t)}{\alpha(t)} = \psi_2(c),$$

where ψ_2 can be either 0 or $\psi_2(c) = D\frac{c^\rho - 1}{\rho}$ for some $D \neq 0$ and $\rho \in \mathbb{R}$. Now note that using Slutsky's theorem we have

$$\begin{aligned}
&\frac{X^*_{pk:k} - X^*_{pk/2:k/2}}{\alpha(n/k)} \\
&= \frac{X^*_{pk:k} - \beta(n/k)}{\alpha(n/k)} - \frac{X^*_{pk/2:k/2} - \beta(2n/k)}{\alpha(2n/k)} \times \frac{\alpha(2n/k)}{\alpha(n/k)} \\
&\quad - \frac{\beta(2n/k) - \beta(n/k)}{\alpha(n/k)} \\
&\xrightarrow{P} H^{\leftarrow}(p) - H^{\leftarrow}(p)2^\rho - \psi_2(2),
\end{aligned}$$

and also,

$$\begin{aligned}
&\frac{X^*_{pk:k} - X^*_{(p/2)k:k}}{\alpha(n/k)} \\
&= \frac{X^*_{pk:k} - \beta(n/k)}{\alpha(n/k)} - \frac{X^*_{(p/2)k:k} - \beta(n/k)}{\alpha(n/k)} \\
&\xrightarrow{P} H^{\leftarrow}(p) - H^{\leftarrow}(p/2).
\end{aligned}$$

Since $H^{\leftarrow}(p) - H^{\leftarrow}(p/2) \neq 0$, another use of Slutsky gives us

$$\text{Pickandsish}_{k,n}(p) = \frac{(X^*_{pk:k} - X^*_{pk/2:k/2})\alpha(n/k)}{(X^*_{pk:k} - X^*_{(p/2)k:k})\alpha(n/k)} \xrightarrow{P} \frac{H^{\leftarrow}(p)(1 - 2^\rho) - \psi_2(2)}{H^{\leftarrow}(p) - H^{\leftarrow}(p/2)}.$$

$\square$

Corollary 3.2.5. *Suppose there exists* $0 < p_1 < p_2 < 1$ *such that* $H^{\leftarrow}(p_1) < H^{\leftarrow}(p_2)$*, and for* $i = 1, 2$*,* $H^{\leftarrow}(p_i) - H^{\leftarrow}(p_i/2) \neq 0$*. Then under the conditions of Proposition 3.2.4,* μ^* *is a product measure if and only if*

$$\textit{Pickandsish}_{k,n}(p_i) \xrightarrow{P} 0, \quad i = 1, 2.$$

Proof. 1^o. Assume that μ^* is a product measure. Then $(\psi_1, \psi_2) \equiv (1, 0)$, i.e., $\rho = 0$ and $\psi_2 \equiv 0$. Hence

$$H^{\leftarrow}(p)(1 - 2^\rho) - \psi_2(2) = H^{\leftarrow}(p)(1 - 1) - 0 = 0.$$

Therefore, provided $0 < p < 1$ and $H^{\leftarrow}(p) - H^{\leftarrow}(p/2) \neq 0$, Proposition 3.2.4 implies $\text{Pickandsish}_{k,n}(p) \xrightarrow{P} 0$.

2^o. Conversely, suppose $p_1 < p_2$ and $\text{Pickandsish}_{k,n}(p_i) \xrightarrow{P} 0$, $i = 1, 2$. Hence

$$H^{\leftarrow}(p_1)(1 - 2^\rho) - \psi_2(2) = 0 \tag{3.2.38}$$

1. Suppose $\rho = 0$ which means $\psi_1 \equiv 1$. Then (3.2.38) implies $\psi_2(2) = 0$ which implies $\psi_2 \equiv 0$. This means μ^* is a product measure.
2. Suppose $\rho \neq 0$ and $\psi_2 \equiv 0$. Then (3.2.38) implies $H^{\leftarrow}(p_i)(1 - 2^\rho) = 0, i = 1, 2$. This implies $H^{\leftarrow}(p_i) = 0, i = 1, 2$, a contradiction to $H^{\leftarrow}(p_1) < H^{\leftarrow}(p_2)$. So this supposition is not possible.
3. Suppose $\rho \neq 0$ and $\psi_2(c) = D\frac{c^\rho - 1}{\rho}$ for $D \neq 0$. Then (3.2.38) implies $(H^{\leftarrow}(p_i) + \frac{D}{\rho})(1 - 2^\rho) = 0, i = 1, 2$. This means $H^{\leftarrow}(p_i) = -\frac{D}{\rho}, i = 1, 2$, a contradiction to $H^{\leftarrow}(p_1) < H^{\leftarrow}(p_2)$. So this supposition is not possible.

Hence we have that μ^* is a product measure if for $p_1 < p_2$ we have $\text{Pickandsish}_{k,n}(p_i) \xrightarrow{P} 0, i = 1, 2$. $\square$

3.2.4 Kendall's Tau, $\rho_\tau(k,n)$

Classically, Kendall's tau statistic (McNeil et al. [2005]) is used to measure the strength of association between two rankings. We use a slightly modified version of the statistic using data pertaining to the k maximum Y-values: $Y_{(1)} \geq \ldots \geq Y_{(k)}$, their concomitants $X_1^*, \ldots, X_k^*$ and the ranks $R_1, \ldots, R_k$ of $X_1^*, \ldots, X_k^*$. The Kendall's tau statistic is

$$\rho_\tau(k,n) := \frac{4}{k(k-1)} \sum_{1 \leq i < j \leq k} \mathbf{1}_{\{R_i < R_j\}} - 1. \tag{3.2.39}$$

This statistic can also be used to show the appropriateness of the CEV model and to decide if the limit measure is a product. We show that under the CEV model $\rho_\tau(k,n)$ as defined in (3.2.39) converges in probability to a limiting constant and when the CEV model holds with a product measure, the limit is 0.

First we prove a lemma on copulas in $[0,1]^2$ which leads to proving convergence for the statistic $\rho_\tau(k,n)$. Recall that a two dimensional copula is any distribution function defined on $[0,1]^2$ with uniform marginals (McNeil et al. [2005]).

Lemma 3.2.6. *Suppose $\{C_\infty, C_n\ n \geq 1\}$ are copulas on $[0,1]^2$, C_∞ is continuous and $C_n \Rightarrow C_\infty$. Then*

$$\int\limits_{[0,1]^2} C_n(u-, v-) dC_n(u,v) \to \int\limits_{[0,1]^2} C_\infty(u,v) dC_\infty(u,v), \qquad (n \to \infty). \tag{3.2.40}$$

Proof. Since $C_n \Rightarrow C_\infty$, the convergence is uniform, that is, we have

$$||C_n - C_\infty|| := \sup_{(u,v)\in[0,1]^2} |C_n(u,v) - C_\infty(u,v)| \to 0.$$

Therefore

$$
\begin{aligned}
&\Big| \int\limits_{[0,1]^2} C_n(u-,v-)dC_n(u,v) - \int\limits_{[0,1]^2} C_\infty(u,v)dC_\infty(u,v)\Big| \\
\le& \int\limits_{[0,1]^2} |C_n(u-,v-) - C_\infty(u,v)|dC_n(u,v) \\
&+ \Big| \int\limits_{[0,1]^2} C_\infty(u,v)dC_n(u,v) - \int\limits_{[0,1]^2} C_\infty(u,v)dC_\infty(u,v)\Big| \\
\le& ||C_n - C_\infty|| + \Big| \int\limits_{[0,1]^2} C_\infty(u,v)dC_n(u,v) - \int\limits_{[0,1]^2} C_\infty(u,v)dC_\infty(u,v)\Big| \\
&\to 0.
\end{aligned}
$$

□

Remark 3.2.1. From Lemma 3.2.6 we get that if $\{C_n; n \ge 1\}$ are random probability measures and C_∞ is continuous, then

$$
\int\limits_{[0,1]^2} C_n(u-,v-)dC_n(u,v) \xrightarrow{P} \int\limits_{[0,1]^2} C_\infty(u,v)dC_\infty(u,v) \tag{3.2.41}
$$

Define the following copulas on $[0,1]^2$:

$$
C_{\mu_n^*}(x,y) := \frac{1}{k}\sum_{i=1}^{k} \epsilon_{\frac{R_i^k}{k},\frac{i}{k}}([0,x]\times[0,y]), \qquad (x,y)\in[0,1]^2 \tag{3.2.42}
$$

$$
C_{\mu^*}(x,y) := \mu^*([\infty, H^{\leftarrow}(x)]\times[y^{-1},\infty]). \tag{3.2.43}
$$

Proposition 3.2.7. *Suppose* $(X_1,Y_1),(X_2,Y_2),\ldots,(X_n,Y_n)$ *are i.i.d. observations from a CEV model which follows* (2.1.2)-(2.1.6) *and suppose* H *defined in* (2.1.6) *is continuous. Assume that* $k = k(n) \to \infty, n \to \infty$ *and* $k/n \to 0$. *Then*

$$
\rho_\tau(k,n) \xrightarrow{P} 4\int\limits_{[0,1]^2} C_{\mu^*}(x,y)dC_{\mu^*}(x,y) - 1 =: J_{\mu^*}. \tag{3.2.44}
$$

If μ^* *is a product measure,* $J_{\mu^*} = 0$

Proof. Proposition 3.2.1 implies that as $k, n \to \infty$ with $k/n \to 0$, for $0 \le x \le 1$, $z \ge 1$

$$\frac{1}{k}\sum_{i=1}^{k} \epsilon_{\left(\frac{R_i}{k}, \frac{k+1}{i}\right)}([0,x] \times [z,\infty]) \Rightarrow \mu^*([-\infty, H^{\leftarrow}(x)] \times (z,\infty]).$$

Therefore, for $0 \le x, y \le 1$,

$$C_{\mu_n^*}(x,y) := \frac{1}{k}\sum_{i=1}^{k} \epsilon_{\left(\frac{R_i}{k}, \frac{i}{k}\right)}([0,x] \times [0,y]) = \frac{1}{k}\sum_{i=1}^{k} \epsilon_{\left(\frac{R_i}{k}, \frac{i}{k}\right)}([0,x] \times [0,y)) + o_P(1)$$
$$\Rightarrow \mu^*([-\infty, H^{\leftarrow}(x)] \times (y^{-1}, \infty]) = C_{\mu^*}(x,y).$$

since H is continuous, and replacing $k+1$ by k does not matter in the limit. This shows that $C_{\mu_n} \Rightarrow C_{\mu^*}$. From Lemma 3.2.6 and Remark 3.2.1 we have

$$S_n^* := \int_0^1\int_0^1 C_{\mu_n^*}(x-, y-) dC_{\mu_n^*}(x,y) \Rightarrow \int_0^1\int_0^1 C_{\mu^*}(x,y) dC_{\mu^*}(x,y).$$

Now note that

$$S_n^* = \int_0^1\int_0^1 C_{\mu_n^*}(x-, y-) dC_{\mu_n^*}(x,y) = \frac{1}{k}\sum_{i=1}^{k} C_{\mu_n^*}\left(\frac{R_i}{k}-, \frac{i}{k}-\right)$$
$$= \frac{1}{k^2}\sum_{i=1}^{k}\sum_{l=1}^{k} \epsilon_{\{\frac{R_l}{k}, \frac{l}{k}\}}([0, \frac{R_i}{k}) \times [0, \frac{i}{k})) = \frac{1}{k^2}\sum_{1 \le l < i \le k} \mathbf{1}_{\{R_l < R_i\}}$$
$$= \frac{k(k-1)}{4k^2}\rho_\tau(k,n) - \frac{1}{k}.$$

Hence we have as $k, n \to \infty$ with $k/n \to 0$,

$$\rho_\tau(k,n) = \frac{k}{k-1}(4S_n^* - 1) + \frac{1}{k-1} \Rightarrow 4\int_0^1\int_0^1 C_{\mu^*}(x,y) dC_{\mu^*}(x,y) - 1 =: J_{\mu^*}.$$

If μ^* is a product, for $0 \le x, y \le 1$,

$$C_{\mu^*}(x,y) := \mu^*([-\infty, H^{\leftarrow}(x)] \times [y^{-1}, \infty]) = y \times H(H^{\leftarrow}(x)) = xy.$$

Hence

$$\int_0^1\int_0^1 C_{\mu^*}(x,y) dC_{\mu^*}(x,y) = \int_0^1\int_0^1 xy\,dx\,dy = \frac{1}{4}$$

and the result follows. $\square$

Proposition 3.2.7 would detect that a limit is not a product if the statistics stabilizes at a non-zero value. We have not been able to prove a limit of 0 implies a product measure and doubt the truth of this statement.

Remark 3.2.2. The three statistics provided above each have their own advantages and disadvantages.

- They are not hard to calculate.
- For the CEV model we have shown that all these statistics stabilize as $k, n \to \infty$ with $k/n \to 0$.
- The rank-based statistics Hillish and Kendall's tau are smooth in nature as the rank transform removes the extremely high or low values.
- The disadvantage of the Pickandsish statistic is that its plot lacks smoothness and exhibits erratic behavior for small data sets.
- Obtaining distributional properties for these statistics would require further limit conditions on the variables, presumably some form of second order behavior.

3.3 Examples and applications

In this section we apply the three estimators proposed in Section 3.2 to data sets and judge their performances in the various cases. First we deal with simulated data from specific models discussed in Section 2.5 of Chapter 2. Then we apply our techniques to Internet traffic data.

3.3.1 Simulation from known CEV limit models

Example 3.3.1. Let X and Y be independent random variables with $X \sim N(0,1)$ and $Y \sim Pareto(1)$. Then the following convergence holds in $\mathbb{M}_+([-\infty,\infty] \times (0,\infty])$ (actually it is an equality)

$$t\mathbf{P}\Big[\Big(X, \frac{Y}{t}\Big) \in [-\infty, x] \times (y,\infty]\Big] = \Phi(x)y^{-1}, \quad -\infty < x < \infty,\ y \geq 1$$

where Φ denotes the standard normal distribution function. We have a CEV model here with $\alpha \equiv 1$, $\beta \equiv 0$, $a(t) = t$, $b \equiv 0$. The limit measure is a product. Hence, theoretically

$$\text{Pickandsish}_{k,n}(p) \xrightarrow{P} 0, 0 < p < 1, \quad \text{Hillish}_{k,n} \xrightarrow{P} 1, \quad \rho_\tau(k,n) \xrightarrow{P} 0.$$

We simulate a sample of size $n = 1000$ and plot the above estimators for $1 \leq k \leq n$. For the Pickandsish statistic we have chosen $p = 0.5$. The simulated data

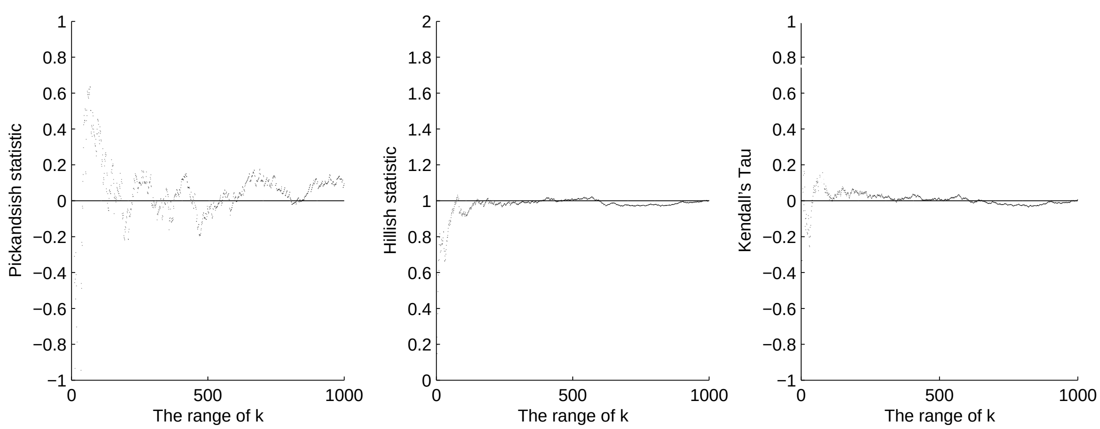

Figure 3.1: $\text{Pickandsish}_{k,n}(0.5)$, $\text{Hillish}_{k,n}$ and $\rho_\tau(k,n)$ for Example 3.3.1

supports the theoretical results stated.

Example 3.3.2. Let X and Z be independent Pareto random variables where $X \sim Pareto(\rho)$ and $Z \sim Pareto(1-\rho)$ with $0 < \rho < 1$. Define $Y = X \wedge Z$. Then we can check that the following holds in $\mathbb{M}_+([0,\infty]\times(0,\infty])$: For $x \geq y > 0$ and t large,

$$t\mathbf{P}\Big[\Big(\frac{X}{t},\frac{Y}{t}\Big)\in[0,x]\times(y,\infty]\Big]=\frac{1}{y^{1-\rho}}\Big(\frac{1}{y^\rho}-\frac{1}{x^\rho}\Big)=\frac{1}{y}\Big(1-\frac{y^\rho}{x^\rho}\Big)=:\mu^*([0,x]\times(y,\infty]).$$

Theoretically the values of the limits of $\text{Hillish}_{k,n}$ and $\text{Pickandsish}_{k,n}(p)$ are as follows.

$$\begin{aligned}
\text{Hillish}_{k,n} &\xrightarrow{P} \int_1^\infty\int_1^\infty \mu^*([0,H^{\leftarrow}(\frac{1}{x})]\times(y,\infty])\frac{dx}{x}\frac{dy}{y}=\int_{x=1}^{\infty}\int_{y=1}^{(\frac{x}{x-1})^{1/\rho}}\frac{1}{y}\Big(1-\frac{y^\rho}{\frac{x}{x-1}}\Big)\frac{dx}{x}\frac{dy}{y}\\
&=\int_{x=1}^{\infty}\Big[1-\frac{1}{1-\rho}\frac{x-1}{x}+\frac{\rho}{1-\rho}(\frac{x-1}{x})^{1/\rho}\Big]\frac{dx}{x}\\
&=\frac{\rho}{1-\rho}\int_{x=1}^{\infty}\sum_{k=2}^{\infty}\binom{1/\rho}{k}\Big(\frac{1}{x}\Big)^{k+1}dx=\frac{\rho}{1-\rho}\sum_{k=2}^{\infty}\binom{1/\rho}{k}\frac{1}{k}.
\end{aligned}$$

Now for $0<p<1$ we have,

$$\text{Pickandsish}_{k,n}(p)\xrightarrow{P}\frac{H^{\leftarrow}(p)(1-2^\rho)-\psi_2(2)}{H^{\leftarrow}(p)-H^{\leftarrow}(p/2)}=\frac{1-2^\rho}{1-\big(\frac{1-p}{1-p/2}\big)^{1/\rho}}.$$

For calculating the Kendall's tau statistics observe that from definition we have:

$$\begin{aligned}
C_{\mu^*}(x,y)&=y\Big(1-\frac{1}{y^\rho}(1-x)\Big)=y-y^{1-\rho}(1-x),\\
dC_{\mu^*}(x,y)&=(1-\rho)y^{-\rho}.
\end{aligned}$$

Hence we have

$$\begin{aligned}
\int_0^1\int_0^1 C_{\mu^*}(x,y)dC_{\mu^*}(x,y)&=(1-\rho)\int_0^1\int_0^1(y-y^{1-\rho}(1-x))y^{-\rho}dxdy\\
&=\frac{1-\rho}{2-\rho}-\frac{1}{4}.
\end{aligned}$$

Therefore

$$\begin{aligned}
\rho_\tau(k,n)&\xrightarrow{P}4\int_0^1\int_0^1 C_{\mu^*}(x,y)dC_{\mu^*}(x,y)-1\\
&=\frac{4(1-\rho)}{2-\rho}-2=-\frac{2\rho}{2-\rho}.
\end{aligned}$$

For $\rho = 0.5$ and $p = 0.5$, theoretically we have

$$\text{Pickandsish}_{k,n}(0.5) \xrightarrow{P} -0.75, \quad \text{Hillish}_{k,n} \xrightarrow{P} 0.5, \quad \rho_\tau(k,n) \xrightarrow{P} -0.67.$$

We simulate a sample of size $n = 1000$ with $\rho = 0.5$ and plot the three statistics for $1 \leq k \leq n$. For the Pickandsish statistic we have chosen $p = 0.5$. The graphs

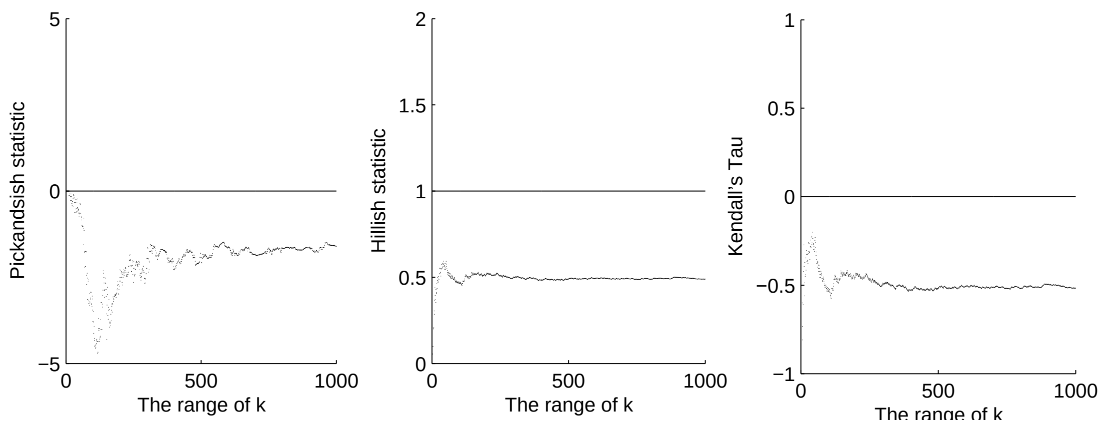

Figure 3.2: Pickandsish$_{k,n}(0.5)$, Hillish$_{k,n}$ and $\rho_\tau(k,n)$ for Example 3.3.2

are consistent with the obtained theoretical limits.

3.3.2 Internet traffic data

The internet forms a network computers where enormous amount of information and resources are exchanged between users. Consider data being transmitted between a pair of internet servers. The natural quantities of interest are the amount of data being transmitted in a session, the transmission duration and the average transmission rate. Empirical evidence has shown that the former two variables are often heavy-tailed in nature [Maulik et al., 2002, Resnick, 2003, Sarvotham et al., 2005].

We study a particular data set of GPS-synchronized traces that were recorded at the University of Auckland *http://wand.cs.waikato.ac.nz/wits*. The data has been downloaded and processed into sessions by Luis Lopez Oliveros, Cornell University.The raw data contains measurements on packet size, arrival time, source and destination IP, port number, Internet protocol, etc. We consider traces corresponding exclusively to incoming TCP traffic sent on December 8, 1999, between 3 and 4 p.m. The packets were clustered into end-to-end (e2e) sessions which are clusters of packets with the same source and destination IP address such that the delay between arrival of two successive packets in a session is at most two seconds. We observe three variables $\{(S_i, L_i, R_i) : 1 \leq i \leq 54353\}$:

$$S_i = \text{size or number of bytes transmitted in a session},$$

$$L_i = \text{duration or length of the session},$$

$$R_i = \frac{S_i}{L_i} \quad \text{or average transfer rate associated with a session.}$$

First, let us look at some summary statistics from the data:

Statistics	S (in bytes)	L (in sec)	R (in bytes/sec)
min	56	0.000009	26.1468
Q1	359	0.5584	305.9999
median	903	1.4155	1019.1
Q3	6000	3.3976	4399.0
max	20566937	1060.2	12591000
Std. dev.	194990	14.8710	276000

The summary shows a concentration towards smaller values and a standard modeling procedure will fail to capture the right tail of the data for S and L.

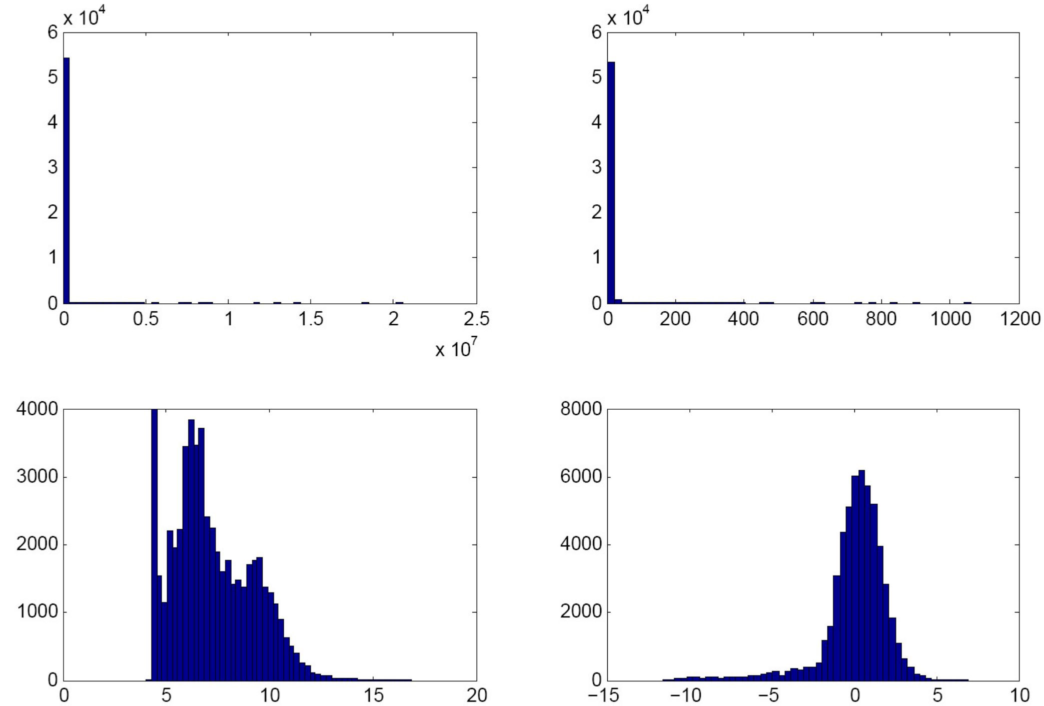

Figure 3.3: Histograms: top row - S and L; bottom row - $\log S$ and $\log L$.

Hence the need for extreme-value modeling. We provide histograms of S, L and $\log S, \log L$ in Figure 3.3 with 50 equi-spaced bins which supports our premise. A bivariate histogram of $\log S$ and $\log L$ with 50×50 bins in Figure 3.4 illuminates this further.

One way to model the data using EVT is to go through the *Peaks Over Threshold* or POT method. Here the data is assumed to follow a Generalized Pareto Distribution (GPD) over a threshold and below the threshold it is either assumed to have some other parametric form or is estimated empirically. A discussion of the POT method can be found in [Embrechts et al., 1997, Chap 6]. Instead of this technique we use the CEV to model our data set as we will see evidence of the transfer rate R, not being in any extreme-value domain of attraction.

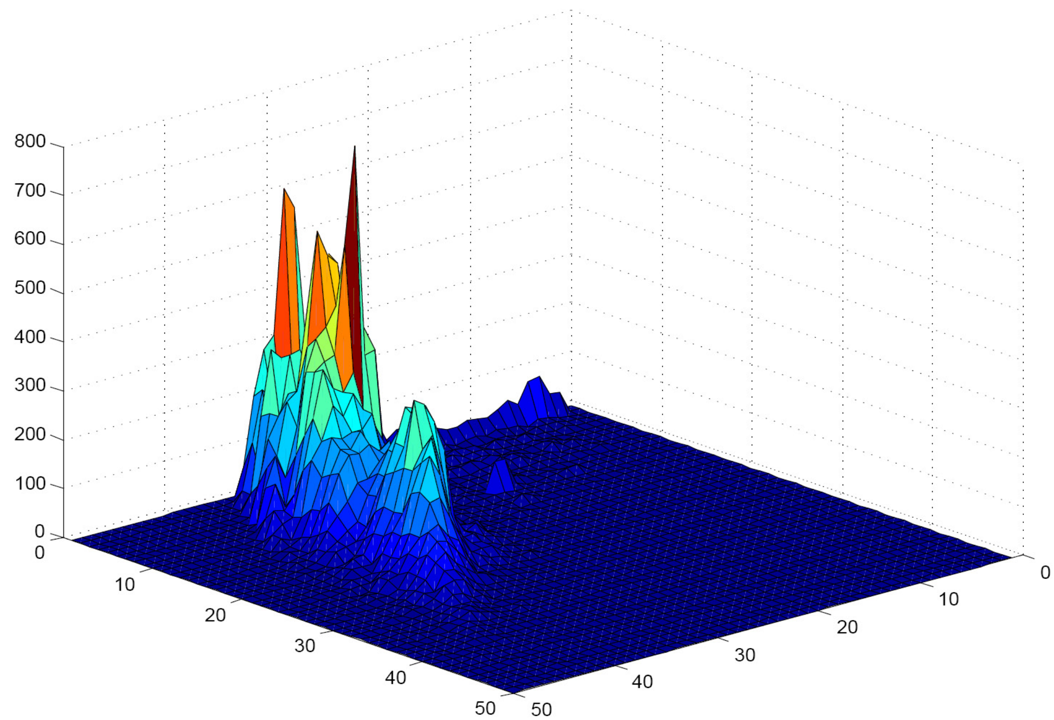

Figure 3.4: Bivariate histogram of $\log S$ and $\log L$.

First we check whether the individual variables are heavy-tailed or not. The Pickands estimator and moment estimators are weakly consistent for the extreme value parameter γ (de Haan and Ferreira [2006]) when the distribution of the variable under consideration is in $D(G_\gamma)$ as in (1.1.1). We plot these estimators over $1 \leq k \leq n$ and observe whether they stabilize over an interval. The Pickands plot indicates that the Pickands estimates of the extreme value parameter are stable for size and duration but not for the transfer rate. The moment plot on the other hand shows that the moment estimate of the extreme-value parameter stabilizes for duration but does not do that clearly for either size or transfer rate. Recall that the CEV model is applicable if either of the variables is in the domain of attraction of an extreme-value distribution. Clearly there is an indication that transfer rate might not be in an extreme value domain.

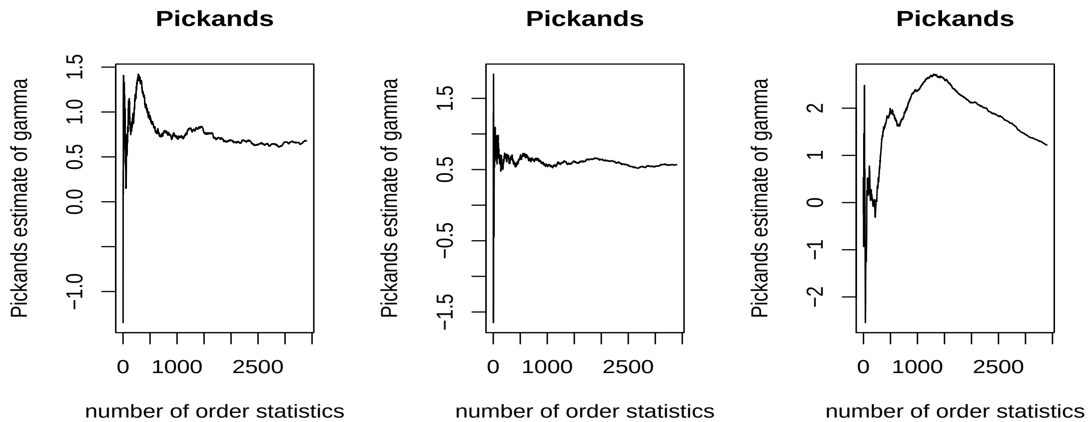

Figure 3.5: Pickands plot of the EV parameter for S, L and R.

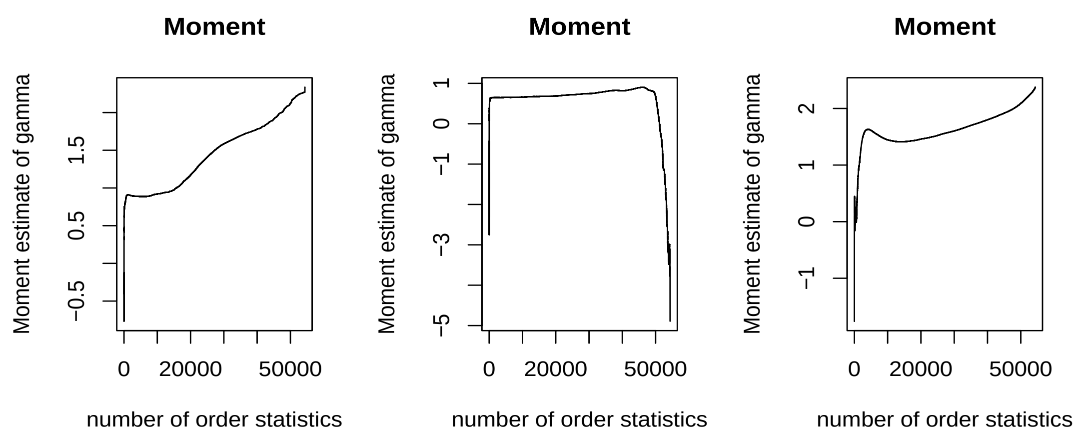

Figure 3.6: Moment estimate plot of the EV parameter for S, L and R.

Now we turn to the three statistics we have devised in this paper first to detect whether we have a CEV model and then to check whether the limit measure is a product. First we consider the pair (R, L) assuming the distribution of L is in $D(G_\gamma)$ for some $\gamma \in \mathbb{R}$. Then we consider the pair (R, S) assuming the distribution of S is in $D(G_\lambda)$ for some $\lambda \in \mathbb{R}$. Observe from Figure 3.7 that neither of the three statistics stabilize for the observations (R, L). Hence a CEV model might not be the right model to apply. On the other hand for (R, S), all

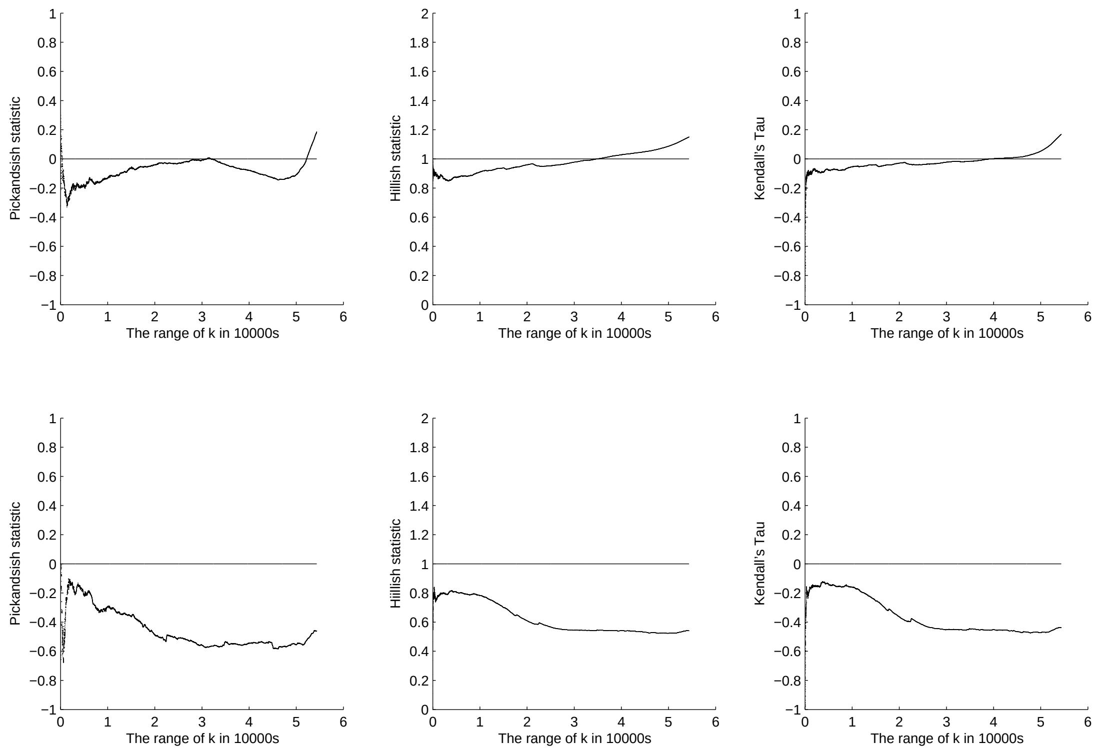

Figure 3.7: The three detectors: top row - R vs. L; bottom row - R vs. S.

the statistics stabilize at some point. But it is clear they are not stabilizing at a point to indicate product measure. Hence we have evidence to model (*transfer rate, size*) as a CEV model with a non-product limit. Note that this also indicates that we should be able to standardize to regular variation on $[0, \infty] \times (0, \infty]$.

The more irregular behavior of the Pickandsish statistic can be attributed to the use of difference of quantiles in the denominator of the statistic which occasionally is very small creating large perturbations. Also note that these are exploratory technique. Our results in Section 3.2 suggest that the statistics converge as $k, n, n/k \to \infty$ but as we have a fixed sample size n here we aim for stability in a two-dimensional plot as k runs from 1 to n.

3.4 Conclusion

The CEV model is intended to provide us with a deeper understanding of multivariate distributions which have some components in an extreme-value domain. In our discussion, we have provided statistics to detect the CEV model in a bivariate set up. These three statistics perform differently for different data sets as we have noted in our examples. A further step would be to find asymptotic distributions for these statistics. On another direction, it would be nice to obtain statistics for detection of conditional models in a multivariate set up of dimension more than two.

CHAPTER 4

QQ PLOTS, RANDOM SETS AND HEAVY TAILS

4.1 Introduction

Given a random sample of univariate data points, a pertinent question is whether this sample comes from some specified distribution F. A variant question is whether the sample is from a location/scale family derived from F. Decision techniques are based on how close the empirical distribution of the sample and the distribution F are for a sample of size n. The empirical distribution function of the i.i.d. random variables $X_1, \ldots, X_n$ is

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq x), \qquad -\infty < x < \infty.$$

The Kolmogorov-Smirnov (KS) statistic is one way to measure the distance between the empirical distribution function and the distribution function F. Glivenko and Cantelli showed (see, for example, Serfling [1980]) that the KS-statistic converges to 0 almost surely. The QQ (or quantile-quantile) plot is another commonly used device to graphically, quickly and informally test the goodness-of-fit of a sample in an exploratory way. It has the advantage of being a graphical tool, which is visually appealing and easy to understand. The QQ plot measures how close the sample quantiles are to the theoretical quantiles. For $0 < p < 1$, the p^{th} *quantile* of F is defined by

$$F^{\leftarrow}(p) := \inf\{x : F(x) \geq p\}. \tag{4.1.1}$$

The sample p^{th} quantile can be similarly defined as $F_n^{\leftarrow}(p)$. If $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics from the sample, then $F_n^{\leftarrow}(p) = X_{\lceil np \rceil:n}$, where as usual $\lceil np \rceil$ is the smallest integer greater than or equal to np. For $0 < p < 1$, $X_{\lceil np \rceil : n}$ is a strongly consistent estimator of $F^{\leftarrow}(p)$ [Serfling, 1980, page 75].

Rather than considering individual quantiles, the QQ plot considers the sample as a whole and plots the sample quantiles against the theoretical quantiles of the specified target distribution F. If we have a correct target distribution, the QQ plot hugs a straight line through the origin at an angle of $45°$. Sometimes we have a location and scale family correctly specified up to unspecified location and scale and in such cases, the QQ plot concentrates around a straight line with some slope (not necessarily $45°$) and intercept (not necessarily 0); the slope and intercept estimate the scale and location. Since a variety of estimation and inferential procedures in the practice of statistics depends on the assumption of normality of the data, the normal QQ plot is one of the most commonly used.

It is intuitive and widely believed that the QQ plot should converge to a straight line as the sample size increases. Our goal here is to formally prove the convergence of the QQ plot considered as a random closed set in $\mathbb{R}^2$. This set of points that form the QQ plot in $\mathbb{R}^2$ is

$$\boldsymbol{S}_n := \{(F^{\leftarrow}(\frac{i}{n+1}), X_{i:n}), \;\; 1 \leq i \leq n\} \tag{4.1.2}$$

where the function $F^{\leftarrow}(\cdot)$ is defined by (4.1.1). For each n, $\boldsymbol{S}_n$ is a random closed set. Note that if $\{\boldsymbol{S}_n\}$ has an almost sure limit $\boldsymbol{S}$, then this limit set must be almost surely constant by the Hewitt-Savage $0-1$ law [Billingsley, 1995]. A straight line (or some closed subset of a straight line) is also a closed set in $\mathbb{R}^2$. Under certain regularity conditions on F, we show that the random set $\boldsymbol{S}_n$ converges in probability to a straight line (or some closed subset of a straight line), in a suitable topology on closed subsets of $\mathbb{R}^2$. We also show the asymptotic consistency of the slope of the least squares line through the QQ plot when the distribution F has bounded support and eventually extend these ideas to the case of heavy-tailed distributions.

Section 4.2 is devoted to preliminary results on the convergence of random closed sets. We also discuss a standard result on convergence of quantiles and, because of our interest in heavy tails, we introduce the concept of regular variation. In Section 4.3, we assume the random variables have a specified distribution F and we consider convergence of the random closed sets $\boldsymbol{S}_n$ forming the QQ plot. In Section 4.4, we discuss how to apply the QQ plot to heavy-tailed data. We assume that the distribution tail is regularly varying with unknown tail index or, slowly varying component and show how the QQ plot can verify the heavy-tailed assumption and estimate α. The usual QQ plot is not informative here in a statistical sense and hence must be modified by thresholding and transformation.

In Example 4.3.1 we have convergence of a log-transformed version of the QQ plot to a straight line when the distribution of the random sample is Pareto. Now Pareto being a special case of a distribution with regularly varying tail, we use the same plotting technique for random variables having a regularly varying tail after thresholding the data. We provide a convergence in probability result considering the $k = k(n)$ upper order statistics of the data set where $k \to \infty$ and $k/n \to 0$. Note that this technique has been used in Chapter 3 in order to detectect heavy-tails marginally in the process of detecting a CEV model. In Section 4.5, a continuity result is provided for a least squares line through these special kinds of closed sets. Section 4.6 connects the ideas of sections 4.4 and 4.5 to give added perspective on the known asymptotic consistency of the slope of the least squares line through the QQ plot as a tail index estimator for the heavy-tailed distribution considered. See Kratz and Resnick [1996], Beirlant et al. [1996].

Notational convention: Bold fonts with small letters are used for vectors, bold fonts with capital letters for sets and calligraphic capital letters for collection of sets.

4.2 Preliminaries

4.2.1 Closed sets and the Fell topology

We denote the distance between the points $\boldsymbol{x}$ and $\boldsymbol{y}$ by $d(\boldsymbol{x}, \boldsymbol{y})$; $\mathcal{F}_d, \mathcal{G}_d$ and $\mathcal{K}_d$ are the classes of closed, open and compact subsets of $\mathbb{R}^d$ respectively. The subscript specifying the dimension of the space is dropped for convenience and used only when this needs to be emphasized for clarity. We are interested in closed sets because the sets of interest such as $\boldsymbol{S}_n$ are random closed sets. There are several ways to define a topology on the space of closed sets. The Vietoris topology and the Fell topology are frequently used and these are hit-or-miss kinds of topologies. We shall discuss the Fell topology below. For further discussion refer to Beer [1993], Matheron [1975], Molchanov [2005].

For a set $\boldsymbol{B} \subset \mathbb{R}^d$, define $\mathcal{F}_B$ as the class of closed sets hitting $\boldsymbol{B}$ and $\mathcal{F}^{\boldsymbol{B}}$ as the class of closed sets disjoint from $\boldsymbol{B}$:

$$\mathcal{F}_{\boldsymbol{B}} = \{\boldsymbol{F} : \boldsymbol{F} \in \mathcal{F},\ \ \boldsymbol{F} \bigcap \boldsymbol{B} \neq \emptyset\}, \quad \mathcal{F}^{\boldsymbol{B}} = \{\boldsymbol{F} : \boldsymbol{F} \in \mathcal{F},\ \ \boldsymbol{F} \bigcap \boldsymbol{B} = \emptyset\}.$$

Now the space $\mathcal{F}$ can be topologized by the Fell topology which has as its sub-base the families $\{\mathcal{F}^{\boldsymbol{K}}, \boldsymbol{K} \in \mathcal{K}\}$ and $\{\mathcal{F}_{\boldsymbol{G}}, \boldsymbol{G} \in \mathcal{G}\}$.

A sequence $\{\boldsymbol{F}_n\}$ converges in the Fell topology towards a limit $\boldsymbol{F}$ in $\mathcal{F}$ (written $\boldsymbol{F}_n \to \boldsymbol{F}$) if and only if it satisfies two conditions:

1. If an open set $\boldsymbol{G}$ hits $\boldsymbol{F}$, $\boldsymbol{G}$ hits all $\boldsymbol{F}_n$, provided n is sufficiently large.
2. If a compact set $\boldsymbol{K}$ is disjoint from $\boldsymbol{F}$, it is disjoint from $\boldsymbol{F}_n$ for all sufficiently large n.

The following result [Matheron, 1975] provides useful conditions for convergence.

Lemma 4.2.1. *For $\boldsymbol{F}_n, \boldsymbol{F} \in \mathcal{F}, n \geq 1, \boldsymbol{F}_n \to \boldsymbol{F}$ as $n \to \infty$ if and only if the following two conditions hold:*

$$\bullet \text{For any } \boldsymbol{y} \in \boldsymbol{F}, \text{ for all large } n, \exists\, \boldsymbol{y}_n \in \boldsymbol{F}_n \text{ s.t. } d(\boldsymbol{y}_n, \boldsymbol{y}) \overset{n\to\infty}{\longrightarrow} 0. \tag{4.2.1}$$

$$\bullet \text{For any subsequence } \{n_k\}, \text{ if } \boldsymbol{y}_{n_k} \in \boldsymbol{F}_{n_k} \text{ converges, then } \lim_{k\to\infty} \boldsymbol{y}_{n_k} \in \boldsymbol{F}. \tag{4.2.2}$$

Furthermore, convergence of sets $\boldsymbol{S}_n \to \boldsymbol{S}$ in $\mathcal{K}$ (with the relativized Fell topology from $\mathcal{F}$) is equivalent to the analogues of (4.2.1) and (4.2.2) holding as well as $\sup_{j\geq 1} \sup\{\|\boldsymbol{x}\| : \boldsymbol{x} \in \boldsymbol{S}_j\} < \infty$ for some norm $\|\cdot\|$ on $\mathbb{R}^d$.

We are going to define random sets in the next subsection. Lemma 4.2.1 can be used to characterize almost sure convergence or convergence in probability of a sequence of random sets to a non-random limit.

The following definition provides a natural and customary notion of distance between compact sets. It will be useful in finding examples.

Definition 4.2.1 (Hausdorff Metric)**.** *Suppose $d : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}_+$ is a metric on $\mathbb{R}^d$ and for $\boldsymbol{S} \in \mathcal{K}$ and $\delta > 0$, define the δ-neighborhood or δ-swelling of S as*

$$\boldsymbol{S}^\delta = \{\boldsymbol{x} : d(\boldsymbol{x}, \boldsymbol{y}) < \delta \ \text{ for } \ \text{ some } \ \boldsymbol{y} \in \boldsymbol{S}\}. \tag{4.2.3}$$

Then for $\boldsymbol{S}, \boldsymbol{T} \in \mathcal{K}$, define the Hausdorff metric [Matheron, 1975] $D : \mathcal{K} \times \mathcal{K} \to \mathbb{R}_+$ by

$$D(\boldsymbol{S}, \boldsymbol{T}) = \inf\{\delta : \boldsymbol{S} \subset \boldsymbol{T}^\delta, \boldsymbol{T} \subset \boldsymbol{S}^\delta\}. \tag{4.2.4}$$

The topology usually used on $\mathcal{K}$ is the *myopic topology* with sub-base elements $\{\mathcal{K}^F, F \in \mathcal{F}\}$ and $\{\mathcal{K}_{\boldsymbol{G}}, \boldsymbol{G} \in \mathcal{G}\}$. The myopic topology on $\mathcal{K}$ is stronger than the Fell topology relativized to $\mathcal{K}$. The topology on $\mathcal{K}' = \mathcal{K} \setminus \{\emptyset\}$ generated by the Hausdorff metric is equivalent to the myopic topology on $\mathcal{K}'$ [Molchanov, 2005, page 405]. Hence convergence in the Hausdorff metric would imply convergence in Fell topology relativized to $\mathcal{K}$ but not the other way round. This idea suffices for our examples.

In certain cases, convergence on $\mathcal{F}$ can be reduced to convergence on $\mathcal{K}$.

Lemma 4.2.2. *Suppose $\boldsymbol{F}_n, \boldsymbol{F} \in \mathcal{F}, n \geq 1$ and there exist $\mathcal{K}_1 \subset \mathcal{K}$ satisfying*

1. $\displaystyle\bigcup_{\boldsymbol{K} \in \mathcal{K}_1} \boldsymbol{K} = \mathbb{R}^d$.
2. *For $\delta > 0$ and $\boldsymbol{K} \in \mathcal{K}$, we have $\overline{\boldsymbol{K}^\delta} \in \mathcal{K}_1$.*
3. $\boldsymbol{F}_n \bigcap \boldsymbol{K} \to \boldsymbol{F} \bigcap \boldsymbol{K}, \quad \forall \boldsymbol{K} \in \mathcal{K}_1$.

Then $\boldsymbol{F}_n \to \boldsymbol{F}$ in $\mathcal{F}$.

Remark 4.2.1. The converse is false. Let $d = 1$, $\boldsymbol{F}_n = \{1/n\}$, $\boldsymbol{F} = \{0\}$ and $\boldsymbol{K} = [-1, 0]$. Then $\boldsymbol{F}_n \to \boldsymbol{F}$ but

$$\boldsymbol{F}_n \bigcap \boldsymbol{K} = \emptyset \not\to \boldsymbol{F} \bigcap \boldsymbol{K} = \boldsymbol{F}.$$

The operation of intersection is not a continuous operation in $\mathcal{F} \times \mathcal{F}$ [Molchanov, 2005, page 400]; it is only upper semicontinuous [Matheron, 1975, page 9].

Proof. We use Lemma 4.2.1 in both directions. If $\boldsymbol{x} \in \boldsymbol{F}$, there exist $\boldsymbol{K} \in \mathcal{K}_1$ and $\boldsymbol{x} \in \boldsymbol{K}$. So $\boldsymbol{x} \in \boldsymbol{F} \cap \boldsymbol{K}$ and from Lemma 4.2.1, since $\boldsymbol{F}_n \cap \boldsymbol{K} \to \boldsymbol{F} \cap \boldsymbol{K}$ as $n \to \infty$, we have existence of $\boldsymbol{x}_n \in \boldsymbol{F}_n \cap \boldsymbol{K}$ and $\boldsymbol{x}_n \to \boldsymbol{x}$. So we have produced $\boldsymbol{x}_n \in \boldsymbol{F}_n$ and $\boldsymbol{x}_n \to \boldsymbol{x}$ as required for (4.2.1).

To verify (4.2.2), suppose $\{\boldsymbol{x}_{n_k}\}$ is a subsequence such that $\boldsymbol{x}_{n_k} \in \boldsymbol{F}_{n_k}$ and $\{\boldsymbol{x}_{n_k}\}$ converges to, say, $\boldsymbol{x}_\infty$. We need to show $\boldsymbol{x}_\infty \in F$. There exists $\boldsymbol{K}_\infty \in \mathcal{K}_1$ such that $\boldsymbol{x}_\infty \in \boldsymbol{K}_\infty$. For any $\delta > 0$, $\boldsymbol{x}_{n_k} \in \overline{\boldsymbol{K}^\delta_\infty} \in \mathcal{K}_1$ for all sufficiently large n_k. So $\boldsymbol{x}_{n_k} \in \boldsymbol{F}_{n_k} \cap \overline{\boldsymbol{K}^\delta_\infty}$. Since $\boldsymbol{F}_{n_k} \cap \overline{\boldsymbol{K}^\delta_\infty} \to \boldsymbol{F} \cap \overline{\boldsymbol{K}^\delta_\infty}$, we have $\lim_{k\to\infty} \boldsymbol{x}_{n_k} = \boldsymbol{x}_\infty \in \boldsymbol{F} \cap \overline{\boldsymbol{K}^\delta_\infty}$. So $\boldsymbol{x} \in \boldsymbol{F}$. □

The next result shows when a point set approximating a curve actually converges to the curve. For this Lemma, $C(0,1]$ is the class of real valued continuous functions on $(0,1]$ and, $D_l(0,1]$ and $D_l(0,\infty]$ are the classes of left continuous functions with finite right hand limits on $(0,1]$ and $(0,\infty]$ respectively.

Lemma 4.2.3. *Suppose* $0 \le x(\cdot) \in C(0,1]$ *is continuous on* $(0,1]$ *and strictly decreasing with* $\lim_{\epsilon\downarrow 0} x(\epsilon) = \infty$. *Suppose further that* $y_n(\cdot) \in D_l(0,1]$ *and* $y(\cdot) \in C(0,1]$ *and* $y_n \to y$ *locally uniformly on* $(0,1]$*; that is, uniformly on compact subintervals bounded away from* 0. *Then for* $k = k(n) \to \infty$,

$$\begin{aligned} &\boldsymbol{F}_n \to \boldsymbol{F} \quad \textit{in } \mathcal{F}_2, \textit{ where} \\ &\boldsymbol{F}_n := \{\big(x(\tfrac{j}{k}), y_n(\tfrac{j}{k})\big); 1 \le j \le k\} \\ &\boldsymbol{F} := \{\big(x(t), y(t)\big); 0 < t \le 1\} = \{\big(u, y(x^{\leftarrow}(u))\big); x(1) \le u < \infty\}. \end{aligned}$$

Proof. Pick $t \in (0,1]$, so that $(x(t), y(t)) \in \boldsymbol{F}$. Then

$$\boldsymbol{F}_n \ni \left(x\Big(\frac{\lceil kt \rceil}{k}\Big), y_n\Big(\frac{\lceil kt \rceil}{k}\Big)\right) \to \big(x(t), y(t)\big) \in \boldsymbol{F},$$

in $\mathbb{R}^2$, verifying (4.2.1). For (4.2.2), Suppose $\left(x\left(\frac{j(n')}{k(n')}\right), y_{n'}\left(\frac{j(n')}{k(n')}\right)\right) \in \boldsymbol{F}_{n'}$ is a convergent subsequence in $\mathbb{R}^2$. Then $x\left(\frac{j(n')}{k(n')}\right)$ is convergent in $\mathbb{R}$ and because $x(\cdot)$ is strictly monotone, $\left(\frac{j(n')}{k(n')}\right)$ converges to some $l \in (0, 1]$. Therefore

$$\boldsymbol{F}_{n'} \ni \left(x\left(\frac{j(n')}{k(n')}\right), y_{n'}\left(\frac{j(n')}{k(n')}\right)\right) \to \big(x(l), y(l)\big) \in \boldsymbol{F},$$

which verifies (4.2.2). $\square$

4.2.2 Random closed sets and weak convergence

In this section, we review definitions and characterizations of weak convergence of random closed sets. In subsequent sections we will show convergence in probability, but since the limit sets will be non-random, weak convergence and convergence in probability coincide. See also Matheron [1975], Molchanov [2005].

Let $(\Omega, \mathcal{A}, P^{'})$ be a complete probability space. $\mathcal{F}$ is the space of all closed sets in $\mathbb{R}^d$ topologized by the Fell topology. Let $\sigma_{\mathcal{F}}$ denote the Borel σ-algebra of subsets of $\mathcal{F}$ generated by this topology. A *random closed set* $\boldsymbol{X} : \Omega \mapsto \mathcal{F}$ is a measurable mapping from $(\Omega, \mathcal{A}, P^{'})$ to $(\mathcal{F}, \sigma_{\mathcal{F}})$. Denote by P the induced probability on $\sigma_{\mathcal{F}}$, that is, $P = P^{'} \circ \boldsymbol{X}^{-1}$. A sequence of random closed sets $\{\boldsymbol{X}_n\}_{n\geq 1}$ weakly converges to a random closed set $\boldsymbol{X}$ with distribution P if the corresponding induced probability measures $\{P_n\}_{n\geq 1}$ converge weakly to P, i.e.,

$$P_n(\mathcal{B}) = P^{'} \circ \boldsymbol{X}_n^{-1}(\mathcal{B}) \to P(\mathcal{B}) = P^{'} \circ \boldsymbol{X}^{-1}(\mathcal{B}), \qquad \text{as } n \to \infty,$$

for each $\mathcal{B} \in \sigma_{\mathcal{F}}$ such that $P(\partial\mathcal{B}) = 0$.

This is not always straightforward to verify from the definition. The following characterization of weak convergence in terms of sup-measures [Vervaat, 1997] is useful. Suppose $h : \mathbb{R}^d \mapsto \mathbb{R}_+ = [0, \infty)$. For $\boldsymbol{X} \subset \mathbb{R}^d$, define $h(\boldsymbol{X}) = \{h(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{X}\}$ and $h^\vee$ is the sup-measure generated by h defined by

$$h^\vee(\boldsymbol{X}) = \sup\{h(\boldsymbol{x}) : \boldsymbol{x} \in \boldsymbol{X}\}$$

[Molchanov, 2005, Vervaat, 1997]. These definitions permit the following characterization [Molchanov, 2005, page 87].

Lemma 4.2.4. *A sequence $\{\boldsymbol{X}_n\}_{n \geq 1}$ of random closed sets converges weakly to a random closed set $\boldsymbol{X}$ if and only if $\mathbb{E}h^\vee(\boldsymbol{X}_n)$ converges to $\mathbb{E}h^\vee(\boldsymbol{X})$ for every non-negative continuous function $h : \mathbb{R}^d \mapsto \mathbb{R}$ with a bounded support.*

4.2.3 Convergence of sample quantiles

The sample quantile is a strongly consistent estimator of the population quantile (Serfling [1980], page 75). The weak consistency of sample quantiles as estimators of population quantiles was shown by Smirnov [1949]; see also [Resnick, 1999, page 179]. We will make use of the Glivenko-Cantelli lemma describing uniform convergence of the sample empirical distribution and also take note of the following quantile estimation result.

Lemma 4.2.5. *Suppose F is strictly increasing at $F^{\leftarrow}(p)$ which means that for all $\epsilon > 0$,*

$$F(F^{\leftarrow}(p - \epsilon)) < p < F(F^{\leftarrow}(p + \epsilon)).$$

Then the p^{th} sample quantile, $X_{\lceil np \rceil : n}$ is a weakly consistent quantile estimator,

$$X_{\lceil np \rceil : n} \xrightarrow{P} F^{\leftarrow}(p)$$

As before, $\lceil np \rceil$ is the 1^{st} integer $\geq np$ and $X_{i:n}$ is the i^{th} smallest order statistic.

4.3 QQ plots from a known distribution: Random sets converging to a constant set

In this section, we will use the results in Section 4.2 to show the convergence of the random closed sets given by (4.1.2) consisting of the points forming the QQ plot to a non-random set in $\mathbb{R}^2$. We will consider the class of distributions which are continuous and strictly increasing on their support. This result will be derived from the easily proven case where we have i.i.d. random variables from a uniform distribution. We are particularly interested in heavy tailed distributions, so we will provide a special case result for the Pareto distribution which is the exemplar of a heavy tailed distribution.

Proposition 4.3.1. *Suppose $X_1, \ldots, X_n$ are i.i.d. with common distribution $F(\cdot)$ and $X_{1:n} \leq X_{2:n} \leq \ldots \leq X_{n:n}$ are the order statistics from this sample. If F is strictly increasing and continuous on its support, then*

$$\boldsymbol{S}_n := \{(F^{\leftarrow}(\frac{i}{n+1}), X_{i:n}); 1 \leq i \leq n\}$$

converges in probability to

$$\boldsymbol{S} := \{(x, x); x \in support(F)\}$$

in $\mathcal{F}_2$.

Proof. We first prove the proposition for the case when F is uniform on $[0, 1]$. More general cases will follow from this case.

Case 1: $F(\cdot)$ **is the Uniform distribution on** $[0,1]$**.**

Denote the order statistics of $Uniform[0,1]$ by $U_{1:n} \le U_{2:n} \le \ldots \le U_{n:n}$. Define

$$\boldsymbol{U}_n := \{(\frac{i}{n+1}, U_{i:n}), \;\; 1 \le i \le n\} \tag{4.3.1}$$

and

$$\boldsymbol{U} := \{(x,x) : 0 \le x \le 1\}. \tag{4.3.2}$$

We show $\boldsymbol{U}_n \stackrel{a.s.}{\to} \boldsymbol{U}$ in $\mathcal{K}_2$ which implies $\boldsymbol{U}_n \stackrel{P}{\to} \boldsymbol{U}$. First note that

$$\sup_{j \ge 1} \sup\{||\boldsymbol{x}|| : \boldsymbol{x} \in \boldsymbol{U}_j\} < 2$$

where $||\cdot||$ is the Euclidean norm in $\mathbb{R}^2$. We apply the convergence criterion given in Lemma 4.2.1 now. The empirical distribution $U_n(x) = n^{-1}\sum_{i=1}^{n} I(U_i \le x)$ converges uniformly for almost all sample paths to x, $0 \le x \le 1$. Without loss of generality suppose this true for all sample paths. Then for all sample paths, the same is true for the inverse process $U_n^{\leftarrow}(p) = U_{\lceil np \rceil:n}$, $0 \le p \le 1$; that is

$$\sup_{0 \le p \le 1} |U_{\lceil np \rceil:n} - p| \to 0, \quad (n \to \infty).$$

Pick $0 \le y \le 1$ and let $\boldsymbol{y} = (y,y) \in \boldsymbol{U}$. For each n, define $\boldsymbol{y}_n$ by

$$\boldsymbol{y}_n = \Big(\frac{\lceil ny \rceil}{n+1}, U_{\lceil ny \rceil:n}\Big), \tag{4.3.3}$$

so that $\boldsymbol{y}_n \in \boldsymbol{U}_n$. Since $|ny - \lceil ny \rceil| \le 1$, $\lceil ny \rceil/(n+1) \to y$ and since $U_{\lceil ny \rceil:n} \to y$, we have $\boldsymbol{y}_n \to (y,y) \in \boldsymbol{U}$. Hence criterion (4.2.1) from Lemma 4.2.1 is satisfied.

Now suppose we have a subsequence $\{n_k\}$ such that $\boldsymbol{y}_{n_k} \in \boldsymbol{U}_{n_k}$ converges. Then $\boldsymbol{y}_{n_k}$ is of the form $\boldsymbol{y}_{n_k} = (i_{n_k}/(n_k+1), U_{i_{n_k}:n_k})$ for some $1 \le i_{n_k} \le n$ and for some $x \in [0,1]$, we have $i_{n_k}/(n_k+1) \to x$ and hence also $i_{n_k}/n_k \to x$. This implies

$$U_{i_{n_k}:n_k} = U_{\lceil n_k \cdot \frac{i_{n_k}}{n_k} \rceil : n_k} \to x,$$

and therefore $\boldsymbol{y}_{n_k} \to (x,x)$ as required for (4.2.2). Therefore $\boldsymbol{U}_n \stackrel{a.s.}{\to} \boldsymbol{U}$ and this concludes consideration of case 1.

Now we analyze the general case where $X_1, \ldots, X_n$ are i.i.d. with common distribution $F(\cdot)$. According to Lemma 4.2.4, we must prove for any non-negative continuous $h : \mathbb{R}^2 \mapsto \mathbb{R}_+$ with compact support that as $n \to \infty$,

$$\mathbb{E}\big(h^{\vee}(\boldsymbol{S}_n)\big) \to \mathbb{E}\big(h^{\vee}(\boldsymbol{S})\big).$$

Since F is continuous, $F(X_1), F(X_2), \ldots, F(X_n)$ are i.i.d. and uniformly distributed on $[0,1]$. Therefore from case 1 we have that

$$\begin{aligned} \boldsymbol{U}_n &:= \{(\frac{i}{n+1}, F(X_{i:n})); 1 \le i \le n\} \\ &\stackrel{d}{=} \{\frac{i}{n+1}, U_{i:n}); 1 \le i \le n\} \\ &\stackrel{a.s.}{\to} \{(x,x); 0 \le x \le 1\} = \boldsymbol{U}. \end{aligned} \tag{4.3.4}$$

in $\mathcal{K}_2$. We now proceed by considering cases which depend on the nature of the support of F. We will need the following identity. For any closed set X, function $f : \mathbb{R}^2 \mapsto \mathbb{R}_+$ and function $\psi : \mathbb{R}^2 \mapsto \mathbb{R}^2$, we have,

$$f^{\vee} \circ \psi(X) = \sup_{t \in \psi(X)} f(t) = \sup_{s \in X} f(\psi(s)) = \sup_{s \in X} f \circ \psi(s) = (f \circ \psi)^{\vee}(X). \tag{4.3.5}$$

It should be noted here that since F is strictly increasing and continuous on its support, $F^{\leftarrow}$ is unique.

Case 2: The support of F is compact, say $[a, b]$.

This implies $F^{\leftarrow}(0) = a,\ F^{\leftarrow}(1) = b$. Define the map $g : [0,1]^2 \mapsto [a,b]^2$ by

$$g(x, y) = (F^{\leftarrow}(x), F^{\leftarrow}(y)).$$

Since F is strictly increasing, observe that $g(\boldsymbol{U}_n)$=$\boldsymbol{S}_n$ and $g(\boldsymbol{U}) = \boldsymbol{S}$. Define $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ as the extension of g to all of $\mathbb{R}^2$:

$$g^*(x, y) = (g_1(x), g_1(y))$$

where

$$g_1(z) = \begin{cases} F^{\leftarrow}(z), & 0 \leq z \leq 1 \\ a(1+z), & -1 \leq z \leq 0 \\ b(2-z), & 1 \leq z \leq 2 \\ 0 & \text{otherwise.} \end{cases}$$

This makes $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ continuous. Since both $\boldsymbol{U}_n$ and $\boldsymbol{U}$ are subsets of $[0,1] \times [0,1]$, we have $g(\boldsymbol{U}_n) = g^*(\boldsymbol{U}_n)$ and $g(\boldsymbol{U}) = g^*(\boldsymbol{U})$. Let f be a non-negative continuous function on $\mathbb{R}^2$ with bounded support and we have, as $n \to \infty$, using (4.3.5),

$$\begin{aligned} \mathbb{E}f^{\vee}(\boldsymbol{S}_n) &= \mathbb{E}f^{\vee}(g(\boldsymbol{U}_n)) = \mathbb{E}f^{\vee}(g^*(\boldsymbol{U}_n)) \\ &= \mathbb{E}(f \circ g^*)^{\vee}(\boldsymbol{U}_n) \to \mathbb{E}(f \circ g^*)^{\vee}(\boldsymbol{S}). \end{aligned}$$

The previous convergence results from $f \circ g^* : \mathbb{R}^2 \mapsto \mathbb{R}_+$ being continuous with bounded support, $\boldsymbol{U}_n \xrightarrow{P} \boldsymbol{U}$, and Lemma 4.2.4. The term to the right of the convergence arrow above equals

$$= \mathbb{E}f^{\vee}(g^*(\boldsymbol{U})) = \mathbb{E}f^{\vee}(g(\boldsymbol{U})) = \mathbb{E}f^{\vee}(\boldsymbol{S}).$$

Therefore $\boldsymbol{S}_n$ converges to $\boldsymbol{S}$ weakly and hence in probability.

Case 3: The support of F is $\mathbb{R} = (-\infty, \infty)$.

Now define $g : (0,1)^2 \mapsto \mathbb{R}^2$ by

$$g(x,y) = (F^{\leftarrow}(x), F^{\leftarrow}(y)).$$

Since F is strictly increasing, $g(\boldsymbol{U}_n) = \boldsymbol{S}_n$ and $g(\boldsymbol{U} \cap (0,1)^2) = \boldsymbol{S}$. Let f be a continuous function with bounded support in $[-M, M]^2$, for some $M > 0$. Extend the definition of g to all of $\mathbb{R}^2$ by defining $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ as

$$g^*(x,y) = (g_1(x), g_1(y)),$$

where

$$g_1(z) = \begin{cases} F^{\leftarrow}(z), & F(-M) \leq z \leq F(M), \\ -M + M\frac{z - F(-M)}{F(-M-1) - F(-M)}, & F(-M-1) \leq z \leq F(-M), \\ M - M\frac{z - F(M)}{F(M+1) - F(M)}, & F(M) \leq z \leq F(M+1) \\ 0 & \text{otherwise} \end{cases}$$

Therefore $g^* : \mathbb{R}^2 \mapsto \mathbb{R}^2$ is continuous. Now note that since $support(f) \subseteq [-M, M]^2$ and $g(x,y) = g^*(x,y)$ for $(x,y) \in [-M,M]^2$, we will have $f \circ g = f \circ g^*$. Therefore

$$\mathbb{E}f^{\vee}(\boldsymbol{S}_n) = \mathbb{E}f^{\vee}(g(\boldsymbol{U}_n)) = \mathbb{E}(f \circ g)^{\vee}(\boldsymbol{U}_n) = \mathbb{E}(f \circ g^*)^{\vee}(\boldsymbol{U}_n) \to \mathbb{E}(f \circ g^*)^{\vee}(\boldsymbol{U}).$$

As with Case 2, the convergence follows from $f \circ g^* : \mathbb{R}^2 \mapsto \mathbb{R}_+$ being continuous with bounded support, $\boldsymbol{U}_n \xrightarrow{P} \boldsymbol{U}$ and Choquet's Theorem (Lemma 4.2.4). The term to the right of the convergence arrow equals

$$= \mathbb{E}(f \circ g)^{\vee}(\boldsymbol{U}) = \mathbb{E}f^{\vee}(g(\boldsymbol{U})) = \mathbb{E}f^{\vee}(\boldsymbol{S}).$$

Therefore $\boldsymbol{S}_n$ converges to $\boldsymbol{S}$ weakly. But since $\boldsymbol{S}$ is a non-random set, this convergence is true also in probability.

Case 4: The support of F is of the form $[a, \infty)$ or $(-\infty, b]$.

This case can be examined in a similar manner as we have done for Cases 2 and 3 by considering each end-point of the interval of support of F according to its nature. $\square$

Example 4.3.1. Here are two examples of the Proposition where the target distribution is of known form. In the first example the distribution is the exponential and in the second, the distribution is the Pareto. The second is reduced to the first by a log-transform.

(a) If F is exponential with parameter $\alpha > 0$, i.e., $F(x) = 1 - e^{-\alpha x}$, $x > 0$, we have

$$\{(-\frac{1}{\alpha}\log(1 - \frac{i}{n+1}), X_{i:n}); 1 \leq i \leq n\} \xrightarrow{P} \{(x, x) : 0 \leq x < \infty\}.$$

(b) If F is Pareto with parameter $\alpha > 0$, i.e., $F(x) = 1 - x^{-\alpha}$, $x > 1$, we have

$$\{(-\log(1 - \frac{i}{n+1}), \log X_{i:n}); 1 \leq i \leq n\} \xrightarrow{P} \{(x, \frac{x}{\alpha}) : 0 \leq x < \infty\}.$$

4.4 QQ plots: Convergence of random sets in the regularly varying case

The classical QQ plot can be graphed only if we know the hypothesized distribution F at least up to location and scale. We extend QQ plotting to the case where the data is from a heavy tailed distribution; this is a semi-parametric assumption which is more general than assuming the target distribution F is known up to location and scale.

We model a one-dimensional heavy-tailed distribution function F by assuming it has a regularly varying tail with some index $-\alpha$, for $\alpha > 0$; that is, if X has distribution F then,

$$P[X > x] = 1 - F(x) = \bar{F}(x) = x^{-\alpha}L(x), \qquad x > 0 \tag{4.4.1}$$

where L is slowly varying. In at least an exploratory context, how can the QQ plot be used to validate this assumption and also to estimate α? (See [Resnick, 2007, page 106].)

Notice that if we take $L \equiv 1$, F turns out to be a Pareto distribution with parameter α. In Example (4.3.1) (a), we have seen that if F has a Pareto distribution with parameter α, then $\boldsymbol{S}_n$ defined as

$$\begin{aligned} \boldsymbol{S}_n &:= \{(-\log(1 - \frac{i}{n+1}), \log X_{i:n});\ 1 \leq i \leq n\} \\ &\xrightarrow{P} \{(x, \frac{x}{\alpha}); 0 \leq x < \infty\} =: \boldsymbol{S}. \end{aligned} \tag{4.4.2}$$

With this in mind, for a general $\bar{F} \in RV_{-\alpha}$, let us define $\boldsymbol{S}_n$ exactly as in (4.4.2). Then we are able to show that $\boldsymbol{S}_n$ converges in probability to the set

$$\boldsymbol{S} = \{(\alpha x, x + \frac{1}{\alpha} \log L(F^{\leftarrow}(1 - e^{-\alpha x})));\ 0 \leq x < \infty\}. \tag{4.4.3}$$

However, since we do not know the slowly varying function $L(\cdot)$, this result is not useful for inference. Estimating α from such a set is not possible unless $L(\cdot)$ is known, nor is it clear how $\boldsymbol{S}_n$ graphically approximating such a set would allow us to validate the model assumption of a regularly varying tail.

Consequently we concentrate on a different asymptotic regime where the asymptotic behavior of the random closed set can be freed from $L(\cdot)$. For a sample of size n from the distribution F with $\bar{F} \in RV_{-\alpha}$, we construct a QQ plot similar to $\boldsymbol{S}_n$ using only the upper $k = k(n)$ upper order statistics of the sample, where we assume $k = o(n)$. We assume that $d_{\mathcal{F}}(\cdot, \cdot)$ is some translation invariant metric on $\mathcal{F}$ which is compatible with the Fell topology. Note Flachsmeyer [1963/1964] characterized the metrizability of the Fell topology and since $\mathbb{R}^d$ is locally compact, Hausdorff and second countable his results apply and allow the conclusion that $\mathcal{F}$ is metrizable under the Fell topology.

For what follows, when $\boldsymbol{A} \in \mathcal{F}_2$, we write $\boldsymbol{A} + (t_1, t_2) = \{\boldsymbol{a} + (t_1, t_2) : \boldsymbol{a} \in \boldsymbol{A}\}$ for the translation of $\boldsymbol{A}$.

Proposition 4.4.1. *Suppose we have a random sample $X_1, X_2, \dots, X_n$ from F where $\bar{F} \in RV_{-\alpha}$ and $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ are the order statistics in decreasing order. Define*

$$\begin{aligned}\boldsymbol{S}_n &= \{(-\log \frac{j}{n+1}, \log X_{(j)}); j = 1, \dots, k\},\\ \boldsymbol{S}_n^{'} &= \{(-\log \frac{j}{k}, \log \frac{X_{(j)}}{X_{(k)}}); 1 \leq j \leq k\}\end{aligned}$$

where $k = k(n) \to \infty$ and $k/n \to 0$ as $n \to \infty$. Also define

$$\begin{aligned}\boldsymbol{T}_n &= \{(x, \frac{x}{\alpha}); x \geq 0\} + (-\log \frac{k}{n+1}, \log X_{(k)}),\\ \boldsymbol{T} &= \{(x, \frac{x}{\alpha}); 0 \leq x < \infty\}.\end{aligned}$$

Then as $n \to \infty$

$$d_{\mathcal{F}}(\boldsymbol{S}_n, \boldsymbol{T}_n) \xrightarrow{P} 0, \qquad \textit{or alternatively,} \qquad \boldsymbol{S}_n' \xrightarrow{P} \boldsymbol{T}.$$

Remark 4.4.1. So after a logarithmic transformation of the data, we make the QQ plot by only comparing the k largest order statistics with the corresponding theoretical exponential distribution quantiles. This produces an asymptotically linear plot of slope $1/\alpha$ starting from the point $(-\log \frac{k}{n+1}, \log X_{(k)})$.

Proof. Note that we can write

$$\begin{aligned}\boldsymbol{S}_n^{'} &= \{(-\log\frac{j}{k}, \log\frac{X_{(j)}}{X_{(k)}}); 1 \le j \le k\} \\ &= \{(-\log t, \log\frac{X_{([kt])}}{X_{(k)}}); t \in \{\frac{1}{k}, \ldots, \frac{k-1}{k}, 1\}\},\end{aligned}$$

and also write $\boldsymbol{T}$ as

$$\boldsymbol{T} = \{(x, \frac{x}{\alpha}); x \ge 0\} = \{(-\log t, -\frac{1}{\alpha}\log t); 0 < t \le 1\},$$

where we put $x = -\log t$. We first show $\boldsymbol{S}_n' \xrightarrow{P} \boldsymbol{T}$.

Referring to Lemma 4.2.3, set

$$x(t) = -\log t, \quad Y_n(t) = \log\frac{X_{(\lceil kt \rceil)}}{X_{(k)}}, \quad y(t) = -\frac{1}{\alpha}\log t, \qquad 0 < t \le 1.$$

From [Resnick, 2007, page 82, equation (4.18)], we have $Y_n \xrightarrow{P} y$, in $D_l(0,1]$, the left continuous functions on $(0,1]$ with finite right limits, metrized by the Skorohod metric. Suppose $\{n''\}$ is a subsequence. There exists a further subsequence $\{n'\} \subset \{n''\}$ such that $Y_{n'} \xrightarrow{a.s.} y$, in $D_l(0,1]$. This convergence is locally uniform because of continuity of y in $(0,1]$. Hence by Lemma 4.2.3, as $n \to \infty$, $\boldsymbol{S}_{n'}' \xrightarrow{a.s.} \boldsymbol{T}$ in $\mathcal{F}$, and therefore $\boldsymbol{S}_n' \xrightarrow{P} \boldsymbol{T}$, in $\mathcal{F}$.

Now observe that with $\boldsymbol{a}_n := (-\log\frac{k}{n+1}, \log X_{(k)})$, we have

$$\begin{aligned}\boldsymbol{S}_n =& \{(-\log\frac{j}{n+1}, \log X_{(j)}); j = 1, \ldots, k\} \\ =& \{(-\log\frac{j}{k}, \log\frac{X_{(j)}}{X_{(k)}}); j = 1, \ldots, k\} + (-\log\frac{k}{n+1}, \log X_{(k)}) \\ =& \boldsymbol{S}_n^{'} + \boldsymbol{a}_n.\end{aligned}$$

Also,

$$\boldsymbol{T}_n = \{(x, \frac{x}{\alpha}); x \ge 0\} + (-\log\frac{k}{n+1}, \log X_{(k)}) = \boldsymbol{T} + \boldsymbol{a}_n.$$

Now, since $d_{\mathcal{F}}(\boldsymbol{S}_n^{'}, \boldsymbol{T}) \xrightarrow{P} 0$, we get

$$d_{\mathcal{F}}(\boldsymbol{S}_n, \boldsymbol{T}_n) = d_{\mathcal{F}}(\boldsymbol{S}_n^{'} + \boldsymbol{a}_n, \boldsymbol{T} + \boldsymbol{a}_n) = d_{\mathcal{F}}(\boldsymbol{S}_n^{'}, \boldsymbol{T}) \xrightarrow{P} 0,$$

as required. □

4.5 Least squares line through a closed set

4.5.1 Convergence of the least squares line

The previous two sections gave results about the convergence of the QQ plot to a straight line in the Fell topology of $\mathcal{F}_2$. It is always of interest to know whether some functional of closed sets is continuous or not and a functional of particular interest is the slope of the least squares line through the points of the QQ plot. The slope of the least squares line is an estimator of scale for location/scale families and this leads to an estimate of the heavy tail index α; see Kratz and Resnick [1996], Beirlant et al. [1996] and [Resnick, 2007, Section 4.6].

Intuition suggests that when a sequence of finite sets converges to a line, the slope of the least squares line should converge to the slope of the limiting line. However there are subtleties which prevent this from being true in general. We need some restriction on the point sets that converge, since otherwise, a sequence of point sets which are essentially linear except for a vanishing bump, may converge to a line but the bump may skew the least squares line sufficiently to prevent the slope from converging; see Example 4.5.1 below.

The following Proposition provides a condition for the continuity property to hold. First define the subclass $\mathcal{F}_{\text{finite or line}} \subset \mathcal{F}_2$ to be the closed sets of $\mathcal{F}_2$

which are either sets of finite cardinality or closed, bounded line segments. These are the only cases of compact sets where it is clear how to define a least squares line. For $\boldsymbol{F} \in \mathcal{F}_{\text{finite or line}}$, the functional LS is defined in the obvious way:

$$LS(\boldsymbol{F}) = \text{slope of the least squares line through the closed set } \boldsymbol{F}$$

For the next proposition, we consider sets $\boldsymbol{F}_n := \{(x_i(n), y_i(n)) : 1 \leq i \leq k_n\}$ of points and write $\bar{x}_n = \sum_{j=1}^{k_n} x_j(n)/k_n$ and $\bar{y}_n = \sum_{j=1}^{k_n} y_j(n)/k_n$. Also, for a finite set $\boldsymbol{S}_n$, $\#\boldsymbol{S}_n$ denotes the cardinality of $\boldsymbol{S}_n$.

Proposition 4.5.1. *Suppose we have a sequence of sets $\boldsymbol{F}_n := \{(x_i(n), y_i(n)) : 1 \leq i \leq k_n\} \in \mathcal{K}_2$, each consisting of k_n points, which converge to a bounded line segment $\boldsymbol{F} \in \mathcal{K}_2$ with slope m where $|m| < \infty$. Then,*

$$LS(\boldsymbol{F}_n) \to LS(\boldsymbol{F}) = m$$

provided $k_n \to \infty$ and there exists $\delta > 0$ such that

$$p_\delta^n := \frac{\#\Big(\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap \boldsymbol{F}_n\Big)}{\#\boldsymbol{F}_n} \to p_\delta \in [0, 1). \qquad (4.5.1)$$

This Proposition gives a condition for the continuity of the slope functional $LS(\cdot)$ when $\{\boldsymbol{F}_n, n \geq 1\}$ and $\boldsymbol{F}$ are bounded sets in $\mathcal{F}_{\text{finite or line}}$. The next Example shows the necessity of condition (4.5.1), which prevents a set of outlier points from skewing the slope of the least squares line.

Example 4.5.1. For $n \geq 1$, define the sets:

$$\boldsymbol{F}_n = \{(\frac{i}{n}, 0), -n \leq i \leq n; (\frac{1}{n}(1 + \frac{j}{2^n}), \frac{1}{n}(1 + \frac{j}{2^n})), 0 \leq j \leq 2^n\}$$

and

$$\boldsymbol{F} = [-1, 1] \times \{0\}.$$

We develop features about this example.

1. For the cardinality of $\boldsymbol{F}_n$ we have

$$\#\boldsymbol{F}_n = k_n = 2^n + 2n + 2.$$

2. We have $\boldsymbol{F}_n \to \boldsymbol{F}$ in $\mathcal{K}_2$. As before, denote the Hausdorff distance between two closed sets in $\mathcal{K}_2$ by $D(\cdot,\cdot)$ and we have $D(\boldsymbol{F}_n, \boldsymbol{F}) < 3/n \to 0$ as $n \to \infty$.

3. Condition (4.5.1) is *not* satisfied. To see this pick any $n \geq 1$ and observe

$$\bar{x}_n = \bar{y}_n = \frac{3(2^n+1)}{2n(2^n+2n+2)} = \frac{3(2^n+1)}{2nk_n} \sim \frac{3}{2n}.$$

Fix $\delta > 0$. For all n so large that $\delta > 1/(2n)$ we have

$$\begin{aligned} &\frac{\#\Big(\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap \boldsymbol{F}_n\Big)}{\#\boldsymbol{F}_n} \\ &\geq \frac{2^n+1}{2^n+2n+2} \to 1 \quad (n \to \infty). \end{aligned}$$

Obviously for this example, $m = LS(\boldsymbol{F}) = 0$. However, if m_n denotes the slope of the least squares line through $\boldsymbol{F}_n$ then we show that $m_n \to 1 \neq 0 = m$. To see this, observe that conventional wisdom yields,

$$m_n = \frac{\sum\limits_{(x_i(n),y_i(n))\in \boldsymbol{F}_n} (y_i(n) - \bar{y}_n)(x_i(n) - \bar{x}_n)}{\sum\limits_{(x_i(n),y_i(n))\in \boldsymbol{F}_n} (x_i(n) - \bar{x}_n)^2}. \tag{4.5.2}$$

For the numerator we have,

$$
\begin{aligned}
&\sum_{(x_i(n),y_i(n))\in \boldsymbol{F}_n} (y_i(n)-\bar{y}_n)(x_i(n)-\bar{x}_n) \\
=&\sum_{(x_i(n),y_i(n))\in \boldsymbol{F}_n} y_i(n)x_i(n) - k_n\bar{y}_n\bar{x}_n \\
=&\sum_{j=0}^{2^n} \frac{1}{n^2}(1+\frac{j}{2^n})^2 - k_n\Big(\frac{3(2^n+1)}{2nk_n}\Big)^2 \\
=&\frac{1}{n^2}\left(\sum_{j=0}^{2^n}\Big(1+\frac{2j}{2^n}+\frac{j^2}{2^{2n}}\Big) - \frac{9}{4k_n}(2^n+1)^2\right) \\
=&\frac{1}{n^2}\left(2\cdot(2^n+1)+\frac{1}{2^{2n}}\sum_{j=0}^{2^n} j^2 - \frac{9}{4k_n}(2^n+1)^2\right)
\end{aligned}
$$

and using the identity $\sum_{j=1}^{N} j^2 = N(N+1)(N+\frac{1}{2})/3 = N(N+1)(2N+1)/6$, we get the above equal to

$$
\begin{aligned}
=&\frac{1}{n^2}\left(2\cdot(2^n+1)+\frac{1}{2^{2n}}\frac{2^n(2^n+1)(2^n+\frac{1}{2})}{3} - \frac{9}{4k_n}(2^n+1)^2\right) \\
=&\frac{2^n+1}{n^2}\Big(2+\frac{2^n+\frac{1}{2}}{3\cdot 2^n} - \frac{9}{4k_n}(2^n+1)\Big) \sim \frac{k_n}{12n^2}.
\end{aligned}
$$

For the denominator, we use the calculation already done for the numerator:

$$
\begin{aligned}
\sum_{(x_i(n),y_i(n))\in \boldsymbol{F}_n} (x_i(n)-\bar{x}_n)^2 &= \sum_{(x_i(n),y_i(n))\in \boldsymbol{F}_n} x_i(n)^2 - k_n(\bar{x}_n)^2 \\
&= \sum_{i=-n}^{n}(\frac{j}{n})^2 + \sum_{j=0}^{2^n}\frac{1}{n^2}(1+\frac{j}{2^n})^2 - k_n\Big(\frac{3(2^n+1)}{2nk_n}\Big)^2 \\
&= \sum_{i=-n}^{n}(\frac{j}{n})^2 + \sum_{(x_i(n),y_i(n))\in \boldsymbol{F}_n} y_i(n)x_i(n) - k_n\bar{y}_n\bar{x}_n \\
&= \frac{2n(n+1)(2n+1)}{6n^2} + \frac{k_n}{12n^2} + o(\frac{k_n}{12n^2}) \\
&= O(n) + \frac{k_n}{12n^2} + o(\frac{k_n}{12n^2}) \sim \frac{k_n}{12n^2}.
\end{aligned}
$$

Combining the asymptotic forms for numerator and denominator with (4.5.2) yields

$$
m_n \sim \frac{k_n/12n^2}{k_n/12n^2} \sim 1, \quad (n\to\infty),
$$

so $m_n \to 1 \neq 0 = m$, as claimed. $\square$

Proof of Proposition 4.5.1. For $(x_i(n), y_i(n)) \in \boldsymbol{F}_n$, we can write

$$y_i(n) = mx_i(n) + z_i(n), \qquad 1 \leq i \leq k_n. \tag{4.5.3}$$

We want to show that $m_n = LS(\boldsymbol{F}_n) \to m = LS(\boldsymbol{F})$, as $n \to \infty$. Fix $\epsilon > 0$. We will provide N such that for $n > N$, we have $|m_n - m| < \epsilon$.

First of all, condition (4.5.1) allows us to fix $\delta > 0$ such that

$$p_\delta^n := p_n = \frac{\#\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap \boldsymbol{F}_n}{\#\boldsymbol{F}_n} \to p < 1.$$

Choose N_1 such that for $n > N_1$, we have $p_n < \frac{1+p}{2}$ or equivalently that $1 - p_n > \frac{1-p}{2}$. For $\eta > 0$ and $\boldsymbol{F} \in \mathcal{K}_2$, recall the definition of the η-swelling of $\boldsymbol{F}$:

$$\boldsymbol{F}^\eta = \{x : d(x, y) < \eta \ \text{ for } \ \text{some } \ y \in \boldsymbol{F}\}. \tag{4.5.4}$$

Since $D(\boldsymbol{F}_n, \boldsymbol{F}) \to 0$ in $\mathcal{K}_2$, we can choose N_2 such that for all $n > N_2$ we have $\boldsymbol{F}_n \subset \boldsymbol{F}^{\epsilon_1}$ where

$$\epsilon_1 := \frac{2\delta\epsilon(1-p)}{4\sqrt{1+m^2}(2 + 2m + \epsilon(1-p))} = \delta_1 \epsilon \frac{(1-p)}{4\sqrt{1+m^2}}$$

and we have set

$$\delta_1 := \frac{\delta}{1 + m + \frac{1}{2}\epsilon(1-p)} < \delta.$$

The choice of δ_1 is designed to ensure that if for some $(x_i(n), y_i(n))$, we have $|x_i(n) - \bar{x}_n| < \delta_1$, then

$$(x_i(n), y_i(n)) \in (\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta).$$

This follows because

$$\begin{aligned}
&|x_i(n) - \bar{x}_n| \vee |y_i(n) - \bar{y}_n| \\
&< \delta_1 + m\delta_1 + 2\epsilon_1\sqrt{1+m^2} \qquad \text{(see figure 4.1(b))} \\
&= \delta_1 + m\delta_1 + 2\frac{\delta_1\epsilon(1-p)}{4\sqrt{1+m^2}}\sqrt{1+m^2} \qquad \text{(using definition of } \epsilon_1\text{)} \\
&= \delta_1(1 + m + \frac{\epsilon(1-p)}{2}) = \delta. \qquad (4.5.5)
\end{aligned}$$

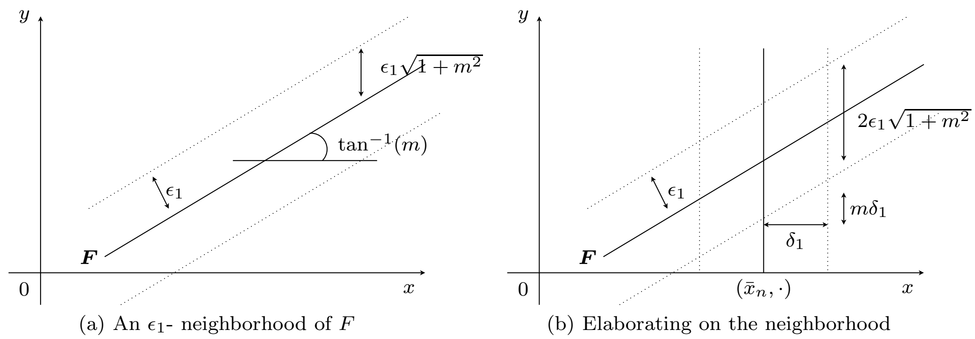

Figure 4.1: Neighborhood of F

Let $N = N_1 \vee N_2$ and restrict attention to $n > N$. Since $\boldsymbol{F}_n \subset \boldsymbol{F}^{\epsilon_1}$, we have for all $1 \leq i \leq k_n$ that $(x_i(n), y_i(n)) \in \boldsymbol{F}^{\epsilon_1}$. By convexity of $\boldsymbol{F}^{\epsilon_1}$, $(\bar{x}_n, \bar{y}_n) \in \boldsymbol{F}^{\epsilon_1}$. Therefore, referring to Figure 4.1(a), we have

$$\begin{aligned}
|z_i(n) - \bar{z}_n| \leq & |y_i(n) - mx_i(n)| + |\bar{y}_n - m\bar{x}_n| \\
\leq & \epsilon_1\sqrt{1+m^2} + \epsilon_1\sqrt{1+m^2} = 2\epsilon_1\sqrt{1+m^2}. \qquad (4.5.6)
\end{aligned}$$

Using the representation (4.5.3) we get,

$$m_n = \frac{\sum_{i=1}^{k_n}(y_i(n) - \bar{y}_n)(x_i(n) - \bar{x}_n)}{\sum_{i=1}^{k_n}(x_i(n) - \bar{x}_n)^2} = m + \frac{\sum_{i=1}^{k_n}(z_i(n) - \bar{z}_n)(x_i(n) - \bar{x}_n)}{\sum_{i=1}^{k_n}(x_i(n) - \bar{x}_n)^2}. \qquad (4.5.7)$$

Therefore,

$$
\begin{aligned}
|m_n - m| &= \left| \frac{\sum_{i=1}^{k_n} (z_i(n) - \bar{z}_n)(x_i(n) - \bar{x}_n)}{\sum_{i=1}^{k_n} (x_i(n) - \bar{x}_n)^2} \right| \\
&\leq \frac{\sum_{i=1}^{k_n} |z_i(n) - \bar{z}_n||x_i(n) - \bar{x}_n|}{\sum_{i=1}^{k_n} (x_i(n) - \bar{x}_n)^2} \\
&\leq 2\epsilon_1 \sqrt{1 + m^2} \frac{\sum_{i=1}^{k_n} |x_i(n) - \bar{x}_n|}{\sum_{i=1}^{k_n} (x_i(n) - \bar{x}_n)^2},
\end{aligned}
$$

where the last inequality follows from (4.5.6). Now for convenience, define the following notations:

$$
\begin{aligned}
|S(x)|_{<\rho} &:= \sum_{|x_i(n) - \bar{x}_n| < \rho} |x_i(n) - \bar{x}_n|, \\
|S(x)|_{\geq\rho} &:= \sum_{|x_i(n) - \bar{x}_n| \geq \rho} |x_i(n) - \bar{x}_n|, \\
S^2(x)_{<\rho} &:= \sum_{|x_i(n) - \bar{x}_n| < \rho} (x_i(n) - \bar{x}_n)^2, \\
S^2(x)_{\geq\rho} &:= \sum_{|x_i(n) - \bar{x}_n| \geq \rho} (x_i(n) - \bar{x}_n)^2, \\
B\big((x, y), \delta\big) &:= (x - \delta, x + \delta) \times (y - \delta, y + \delta).
\end{aligned}
$$

Therefore

$$\begin{aligned}\frac{\sum_{i=1}^{k_n}|x_i(n)-\bar{x}_n|}{\sum_{i=1}^{k_n}(x_i(n)-\bar{x}_n)^2} &= \frac{|S(x)|_{<\delta_1}+|S(x)|_{\geq\delta_1}}{S^2(x)_{<\delta_1}+S^2(x)_{\geq\delta_1}}\\
=&\frac{(|S(x)|_{<\delta_1}+|S(x)|_{\geq\delta_1})/S^2(x)_{\geq\delta_1}}{(S^2(x)_{<\delta_1}/S^2(x)_{\geq\delta_1}+1)}\\
\leq&\frac{(|S(x)|_{<\delta_1}+|S(x)|_{\geq\delta_1})}{S^2(x)_{\geq\delta_1}}\\
\leq&\frac{1}{\delta_1}\left(\frac{|S(x/\delta_1)|_{<1}}{S^2(x/\delta_1)_{\geq 1}}+1\right)\\
\leq&\frac{1}{\delta_1}\left(\frac{\#\{(x_i(n),y_i(n))\in\boldsymbol{F}_n:|x_i(n)-\bar{x}_n|<\delta_1\}}{\#\{(x_i(n),y_i(n))\in\boldsymbol{F}_n:|x_i(n)-\bar{x}_n|\geq\delta_1\}}+1\right)\\
\leq&\frac{1}{\delta_1}\left(\frac{\#\{(x_i(n),y_i(n))\in\boldsymbol{F}_n:(x_i(n),y_i(n))\in B((\bar{x}_n,\bar{y}_n),\delta)\}}{\#\{(x_i(n),y_i(n))\in\boldsymbol{F}_n:(x_i(n),y_i(n))\notin B((\bar{x}_n,\bar{y}_n),\delta)\}}+1\right).\end{aligned}$$

The choice of δ_1 justifies the previous step by (4.5.5). The previous expression is bounded by

$$\leq\frac{1}{\delta_1}(\frac{p_n}{1-p_n}+1)\leq\frac{1}{\delta_1}(\frac{1+p}{1-p}+1)=\frac{2}{\delta_1(1-p)},$$

and we recall $p<1$. Consequently

$$\begin{aligned}|m_n-m| =&2\epsilon_1\sqrt{1+m^2}\frac{\sum_{i=1}^{k_n}|x_i(n)-\bar{x}_n|}{\sum_{i=1}^{k_n}(x_i(n)-\bar{x}_n)^2}\leq 2\epsilon_1\sqrt{1+m^2}\times\frac{2}{\delta_1(1-p)}\\
=&2\epsilon\delta_1\frac{1-p}{4\sqrt{1+m^2}}\sqrt{1+m^2}\times\frac{2}{\delta_1(1-p)}=\epsilon.\end{aligned}$$

This completes the proof that $m_n\to m$ under condition (4.5.1). $\square$

Corollary 4.5.2. *If $\bar{x}_n\to\mu_x<\infty$ and $\bar{y}_n\to\mu_y<\infty$, as $n\to\infty$, then Proposition 4.5.1 holds if we replace $(\bar{x}_n,\bar{y}_n)$ in (4.5.1) by (μ_x,μ_y).*

So in place of condition (4.5.1) we are assuming that there exists $\delta>0$ such that

$$p_\delta^n:=\frac{\#\{(\mu_x-\delta,\mu_x+\delta)\times(\mu_y-\delta,\mu_y+\delta)\}\bigcap\boldsymbol{F}_n}{\#\boldsymbol{F}_n}\to p_\delta\in[0,1). \qquad (4.5.8)$$

Proof. Let us fix $\delta > 0$ such that

$$p_n^* := \frac{\#\{(\mu_x - 2\delta, \mu_x + 2\delta) \times (\mu_y - 2\delta, \mu_y + 2\delta)\} \bigcap \boldsymbol{F}_n}{\#\boldsymbol{F}_n} \to p \in [0, 1).$$

Since $\bar{x}_n \to \mu_x < \infty$ and $\bar{y}_n \to \mu_y < \infty$, there exists N^* such that $n > N^*$ implies that $(\bar{x}_n, \bar{y}_n) \in (\mu_x - \delta, \mu_x + \delta) \times (\mu_y - \delta, \mu_y + \delta)$. Hence for $n > N^*$

$$\begin{aligned} p_n &:= \frac{\#\{(\bar{x}_n - \delta, \bar{x}_n + \delta) \times (\bar{y}_n - \delta, \bar{y}_n + \delta)\} \bigcap \boldsymbol{F}_n}{\#\boldsymbol{F}_n} \\ &\leq \frac{\#\{(\mu_x - 2\delta, \mu_x + 2\delta) \times (\mu_y - 2\delta, \mu_y + 2\delta)\} \bigcap \boldsymbol{F}_n}{\#\boldsymbol{F}_n} \\ &= p_n^* \to p \in [0, 1). \end{aligned}$$

Now choose $N_1 \geq N^*$ such that for all $n > N_1$, we have $p_n < \frac{1+p}{2}$. This also means that $1 - p_n > \frac{1-p}{2}$.

The rest of the proof is the same as that of Proposition 4.5.1. □

4.6 Slope of the LS line as a tail index estimator

For heavy tailed distributions, the slope of the least squares line through the QQ plot made by the upper k_n largest order statistics is a consistent estimator of $1/\alpha$. See Kratz and Resnick [1996], Beirlant et al. [1996] and [Resnick, 2007, Section 4.6]. It should be noted here that this result does not come as a direct consequence of Proposition 4.5.1 which requires the target set to be bounded. Additional work is necessary for such a result to hold and thus we connect the ideas of the previous section with this result.

Proposition 4.6.1. *Consider non-negative random variables $X_1, \dots, X_n$ which are i.i.d with common distribution F where $\bar{F} \in RV_{-\alpha}$ and $X_{(1)} \geq X_{(2)} \geq \dots \geq X_{(n)}$ are the order statistics in decreasing order. As in Proposition 4.4.1, the sets $\boldsymbol{S}_n$ and $\boldsymbol{T}_n$ are*

$$\boldsymbol{S}_n = \{(-\log \frac{j}{n+1}, \log X_{(j)}); j = 1, \dots, k\}$$
$$\boldsymbol{T}_n = \{(x, \frac{x}{\alpha}); x \geq 0\} + (-\log \frac{k}{n+1}, \log X_{(k)}).$$

Write

$$\begin{aligned} \boldsymbol{S}_n^{'} &= \boldsymbol{S}_n + a_n \\ &= \{(-\log \frac{j}{k_n}, \log \frac{X_{(j)}}{X_{(k_n)}}); j = 1, \dots, k_n\} \\ &=: \{(x_j(n), y_j(n)); j = 1, \dots, k_n\}, \quad \textit{and}, \\ \boldsymbol{T} &= \boldsymbol{T}_n + a_n = \{(x, \frac{x}{\alpha}); x \geq 0\} \end{aligned}$$

where $a_n = (\log \frac{k}{n+1}, -\log X_{(k)})$ is a random point. Then,

$$LS(\boldsymbol{S}_n^{'}) = LS(\boldsymbol{S}_n) \xrightarrow{P} \frac{1}{\alpha} = LS(\boldsymbol{T}_n) = LS(\boldsymbol{T}) \tag{4.6.1}$$

as $k := k_n \to \infty$ and $k_n/n \to 0$ as $n \to \infty$.

The result is believable based on the fact that $d_{\mathcal{F}}(\boldsymbol{S}_n, \boldsymbol{T}_n) \xrightarrow{P} 0$ (from Proposition 4.4.1). However, since neither $\boldsymbol{T}_n$ nor $\boldsymbol{T}$ are $\mathcal{K}_2$ sets, some sort of truncation to compact regions of $\mathbb{R}^2$ is necessary in order to capitalize on Proposition 4.5.1. To truncate $\boldsymbol{S}_n^{'}$ and $\boldsymbol{T}$, define for some integer $M > 2$, define $\boldsymbol{K}_M = [0, M] \times [0, 2M/\alpha]$, and let

$$\boldsymbol{S}_n^{'M} = \boldsymbol{S}_n^{'} \cap \boldsymbol{K}_M \quad \text{and} \quad \boldsymbol{T}^M = \boldsymbol{T} \cap \boldsymbol{K}_M.$$

Proof. Some preliminary observations: Clearly, $LS(\boldsymbol{S}_n) = LS(\mathcal{S}_n^{'} + a_n) = LS(\mathcal{S}_n^{'})$ and with $x_j(n), y_j(n)$ defined in the statement of the Proposition,

$$LS(\mathcal{S}_n^{'}) = \frac{\bar{S}_{XY} - \bar{S}_X \bar{S}_Y}{\bar{S}_{XX} - (\bar{S}_X)^2},$$

where, as usual,

$$\bar{S}_X = \frac{1}{k_n} \sum_{(x_j(n), y_j(n)) \in \mathcal{S}'_n} x_j(n), \qquad \bar{S}_Y = \frac{1}{k_n} \sum_{(x_j(n), y_j(n)) \in \mathcal{S}'_n} y_j(n),$$
$$\bar{S}_{XY} = \frac{1}{k_n} \sum_{(x_j(n), y_j(n)) \in \mathcal{S}'_n} x_j(n) y_j(n), \qquad \bar{S}_{XX} = \frac{1}{k_n} \sum_{(x_j(n), x_j(n)) \in \mathcal{S}'_n} (x_j(n))^2.$$

We need similar quantities $\bar{S}_X^M, \bar{S}_Y^M, \bar{S}_{XY}^M$ corresponding to averages of points restricted to $\boldsymbol{K}_M$, so for instance

$$\bar{S}_X^M = \frac{1}{k^M} \sum_{(x_j(n), y_j(n)) \in {\mathcal{S}'_n}^M} x_j(n)$$

and $k^M = \# {\mathcal{S}'_n}^M$. A simple calculation given in [Resnick, 2007, page 109] yields as $k \to \infty$,

$$\bar{S}_X = \frac{1}{k} \sum_{i=1}^{k} (-\log \frac{i}{k}) \sim \int_0^1 (-\log x) dx = 1 \quad \text{and}$$
$$\bar{S}_{XX} = \frac{1}{k} \sum_{i=1}^{k} (-\log \frac{i}{k})^2 \sim \int_0^1 (-\log x)^2 dx = 2, \tag{4.6.2}$$

while for $\bar{S}_Y$ we have

$$\bar{S}_Y = \frac{1}{k} \sum_{i=1}^{k} (-\log \frac{X_{(i)}}{X_{(k)}}) \xrightarrow{P} \frac{1}{\alpha} \tag{4.6.3}$$

since $\bar{S}_Y$ is the Hill estimator and is consistent for $1/\alpha$ [Resnick, 2007, Csörgő et al., 1985, Mason, 1982, Mason and Turova, 1994].

We need the corresponding limits for $\bar{S}_X^M, \bar{S}_{XX}^M, \bar{S}_Y^M$. These calculations and subsequent calculations are simplified by the following facts:

1. The ratios of order statistics process converges, as $k \to \infty$, $k/n \to 0$,

$$\frac{X_{(\lceil kt \rceil)}}{X_{(k)}} \xrightarrow{P} t^{-1/\alpha}, \tag{4.6.4}$$

   in $D_l(0, \infty]$ [Resnick, 2007, page 82].

2. Define the random measure

$$\hat{\nu}_n(\cdot) = \frac{1}{k}\sum_{i=1}^{n} \epsilon_{X_{(i)}/X_{(k)}}(\cdot)$$

on $(0,\infty]$, which puts mass $1/k$ at the points $\{X_{(i)}/X_{(k)}, 1 \le i \le n\}$. Then

$$\hat{\nu}_n \xrightarrow{P} \nu_\alpha, \tag{4.6.5}$$

in the space of Radon measures on $(0,\infty]$, where $\nu_\alpha(x,\infty] = x^{-\alpha}$, $x > 0$ [Resnick, 2007, page 83].

3. The number of points k^M in ${\mathcal{S}'_n}^M$ satisfies, as $n \to \infty$, $k \to \infty$, $k/n \to 0$,

$$k^M/k \xrightarrow{P} 1 - e^{-M}. \tag{4.6.6}$$

To see this, observe

$$\begin{aligned}
k^M/k =& \frac{1}{k}\#\{j \le k : k \ge j \ge ke^{-M} \text{ and } \frac{X_{(j)}}{X_{(k)}} \le e^{2M/\alpha}\} \\
=& \frac{1}{k}\#\{j \le k : 1 \le \frac{X_{(j)}}{X_{(k)}} \le \frac{X_{(\lceil ke^{-M}\rceil)}}{X_{(k)}} \wedge e^{2M/\alpha}\} \\
=& \hat{\nu}_n\Big(1, \frac{X_{(\lceil ke^{-M}\rceil)}}{X_{(k)}} \wedge e^{2M/\alpha}\Big] \\
\xrightarrow{P}& 1 - \Big((e^{-M})^{-1/\alpha} \wedge e^{2M/\alpha}\Big)^{-\alpha} = 1 - e^{-M}.
\end{aligned}$$

We continue using these three facts. For $\bar{S}_X^M$ we have

$$\bar{S}_X^M = \frac{1}{k^M} \sum_{\left(x_i(n), y_i(n)\right) \in {\mathcal{S}'_n}^M} x_i(n) = \frac{1}{k^M} \sum_{\substack{j : k \ge j \ge ke^{-M} \\ 0 < \log X_{(j)}/X_{(k)} \le 2M/\alpha}} -\log\frac{j}{k}.$$

Set

$$\begin{aligned}
\left(\bar{S}_X^M\right)^* :=& \frac{1}{k^M} \sum_{j : k \ge j \ge ke^{-M}} -\log\frac{j}{k} = \frac{k}{k^M}\frac{1}{k} \sum_{j : k \ge j \ge ke^{-M}} -\log\frac{j}{k} \\
\sim& \frac{1}{1-e^{-M}} \int_{e^{-M}}^{1} -\log x \, dx = \frac{1}{1-e^{-M}} \int_0^M ye^{-y} dy \\
=:& 1 + \epsilon_X(M),
\end{aligned}$$

where $\epsilon_X(M) \to 0$ as $M \to \infty$. Also, $\bar{S}_X^M$ and $\left(\bar{S}_X^M\right)^*$ are close asymptotically since

$$P[\bar{S}_X^M \neq \left(\bar{S}_X^M\right)^*] = P\Big\{\bigcup_{k \geq j \geq k^{-M}} [\log \frac{X_{(j)}}{X_{(k)}} > 2M/\alpha]\Big\}$$
$$= P[\log \frac{X_{(\lceil ke^{-M}\rceil)}}{X_{(k)}} > 2M/\alpha] \to 0,$$

since

$$\frac{X_{(\lceil ke^{-M}\rceil)}}{X_{(k)}} \xrightarrow{P} e^{M/\alpha} < e^{2M/\alpha}.$$

We conclude

$$\bar{S}_X^M \xrightarrow{P} 1 + \epsilon_X(M) := \mu_X^M, \tag{4.6.7}$$

with $\epsilon_X(M) \to 0$ as $M \to \infty$, and in a similar way we can derive that

$$\bar{S}_{XX}^M \xrightarrow{P} 2 + \epsilon_{XX}(M), \tag{4.6.8}$$

where $\epsilon_{XX}(M) \to 0$ as $M \to \infty$. For $\bar{S}_Y^M$ we have

$$\begin{aligned}
\bar{S}_Y^M &= \frac{1}{k^M} \sum_{\substack{j: k \geq j \geq ke^{-M} \\ 0 < \log X_{(j)}/X_{(k)} \leq 2\alpha^{-1}M}} \log \frac{X_{(j)}}{X_{(k)}} \\
&= \frac{1}{k^M} \sum_{j: 0 < \log X_{(j)}/X_{(k)} \leq 2\alpha^{-1}M \wedge \log X_{(\lceil ke^{-M}\rceil)}/X_{(j)}} \log \frac{X_{(j)}}{X_{(k)}} \\
&= \frac{k}{k^M} \int_1^{2\alpha^{-1}M \wedge \log X_{(\lceil ke^{-M}\rceil)}/X_{(j)}} \log y \, \hat{\nu}_n(dy) \\
&\xrightarrow{P} \frac{1}{1-e^{-M}} \int_1^{2\alpha^{-1}M \wedge \alpha^{-1}M} \log y \, \nu_\alpha(dy) \\
&= \frac{1}{1-e^{-M}} \int_0^{M/\alpha} s e^{-\alpha s} ds =: \mu_Y^M,
\end{aligned}$$

where $\mu_Y^M \to \frac{1}{\alpha}$ as $M \to \infty$. We conclude

$$\bar{S}_Y^M \xrightarrow{P} \mu_Y^M. \tag{4.6.9}$$

To prove (4.6.1), we follow the following outline of steps.

- Step 1: Prove ${\mathcal{S}'_n}^M \xrightarrow{P} \boldsymbol{T}^M$.

- Step 2: Verify that Corollary 4.5.2 is applicable by showing that the analogue of (4.5.1) holds. This permits the conclusion that

  $$LS({\mathcal{S}'_n}^M) \xrightarrow{P} 1/\alpha.$$

  Coupled with (4.6.7), (4.6.8) and (4.6.9), this yields

  $$\bar{S}^M_{XY} = \frac{2}{\alpha} + \epsilon_{XY}(M) + o_p(1), \tag{4.6.10}$$

  where $\lim_{M\to\infty} \epsilon_{XY}(M) = 0$ and $o_p(1) \xrightarrow{P} 0$ as $n \to \infty$.

- Step 3: Compare $\bar{S}_{XY}$ and $\bar{S}^M_{XY}$ and check that

  $$\lim_{M\to\infty} \limsup_{n\to\infty} P[|\bar{S}^M_{XY} - \bar{S}_{XY}| > \eta] = 0, \quad \forall \eta > 0. \tag{4.6.11}$$

  This gives $\bar{S}_{XY} \xrightarrow{P} 2/\alpha$ which coupled with (4.6.2) and (4.6.3) implies (4.6.1).

We may check Step 1 using a very minor modification of Lemma 4.2.3, following the pattern of proof used for Proposition 4.4.1. For Step 2, the challenge is to verify condition (4.5.8) holds and we defer this to the end of the proof. Thus we turn to Step 3.

First of all, we observe that $\bar{S}^M_{XY}$ and $\bar{S}_{XY}$ average, respectively k^M and k terms but there is no need to differentiate: For any $\eta > 0$,

$$\begin{aligned} H_{M,n} &:= P\Big[\Big|\frac{1}{k^M} \sum_{(x_j(n),y_j(n))\in {\mathcal{S}'_n}^M} x_i(n)y_i(n) - \frac{1}{k} \sum_{(x_j(n),y_j(n))\in {\mathcal{S}'_n}^M} x_i(n)y_i(n)\Big| > \eta\Big] \\ &= P\Big[\Big|\frac{1}{k^M} - \frac{1}{k}\Big| \sum_{(x_j(n),y_j(n))\in {\mathcal{S}'_n}^M} x_i(n)y_i(n) > \eta\Big] \end{aligned}$$

and dividing the sum by k^M yields

$$= P\Big[\bar{S}_{XY}^M\Big|1 - \frac{k^M}{k}\Big| > \eta\Big].$$

Since $\bar{S}_{XY}^M$ is convergent in probability, it is stochastically bounded and since, as $n \to \infty$,

$$\Big|1 - \frac{k^M}{k}\Big| \xrightarrow{P} 1 - (1 - e^{-M}) = e^{-M} \overset{M\to\infty}{\to} 0,$$

we conclude

$$\lim_{M\to\infty} \limsup_{n\to\infty} H_{M,n} = 0. \tag{4.6.12}$$

Next observe for $\eta > 0$,

$$\begin{aligned}
&P[\Big|\frac{1}{k} \sum_{(x_j(n),y_j(n))\in \mathcal{S}'^{\,M}_n} x_j(n)y_j(n) - \frac{1}{k}\sum_{k\geq j\geq ke^{-M}} x_j(n)y_j(n)\Big| > \eta] \\
\leq &P\{\bigcup_{k\geq j\geq ke^{-M}} [\frac{X_{(j)}}{X_{(k)}} > e^{2M/\alpha}]\} \\
\leq &P[\frac{X_{(\lceil ke^{-M}\rceil)}}{X_{(k)}} > e^{2M/\alpha}] \to 0, \quad (n\to\infty).
\end{aligned} \tag{4.6.13}$$

Note that by the Cauchy-Schwartz inequality,

$$\begin{aligned}
&\big(|\bar{S}_{XY} - \frac{1}{k}\sum_{k\geq j\geq ke^{-M}} x_j(n)y_j(n)|\big)^2 \\
\leq &\big(\frac{1}{k}\sum_{1\leq j\leq ke^{-M}} x_j(n)y_j(n)\big)^2 \\
\leq &\frac{1}{k}\sum_{1\leq j\leq ke^{-M}} x_j(n)^2 \;\cdot\; \frac{1}{k}\sum_{1\leq j\leq ke^{-M}} y_j(n)^2.
\end{aligned}$$

Furthermore

$$\frac{1}{k}\sum_{1\leq j\leq ke^{-M}} y_j(n)^2 = \int_{\log X_{(\lceil ke^{-M}\rceil)}/X_{(k)}}^{\infty} (\log y)^2 \hat{\nu}_n(dy)$$

and using (4.6.4), we have for some $c > 0$, all large n and some M that the above is bounded by

$$\int_{cM}^{\infty} (\log y)^2 \hat{\nu}_n(dy) + o_p(1). \tag{4.6.14}$$

Assessing (4.6.12), (4.6.13) and (4.6.14), we see that (4.6.11) will be proved if we show

$$\lim_{M\to\infty}\limsup_{n\to\infty} P\Big[\int_M^\infty (\log y)^2 \hat\nu_n(dy) > \eta\Big] = 0, \quad (\forall \eta > 0). \tag{4.6.15}$$

This treatment is similar to the stochastic version of Karamata's theorem (Feigin and Resnick [1997], [Resnick, 2007, page 207]. For $0 < \zeta < 1 \wedge \alpha$ and large M, the integrand $(\log y)^2$ is dominated by y^ζ. Bound the integral by

$$\int_M^\infty \hat\nu_n(y,\infty]\,\zeta y^{\zeta-1}dy + M^\zeta \hat\nu_n(M,\infty].$$

If we let first $n\to\infty$ and then $M\to\infty$, for the second piece we have

$$M^\zeta\hat\nu_n(M,\infty] \xrightarrow{P} M^\zeta \nu_\alpha(M,\infty] = M^{\zeta-\alpha} \to 0.$$

Now we deal with the integral. Set $b(t) = (1/(1-F))^{\leftarrow}(t)$ so that $X_{(k)}/b(n/k) \xrightarrow{P} 1$ [Resnick, 2007, page 81]. For $\gamma > 0$,

$$\begin{aligned}
&P[\int_M^\infty \hat\nu_n(y,\infty]\,\zeta y^{\zeta-1}dy > \eta]\\
=&P[\int_M^\infty \hat\nu_n(y,\infty]\,\zeta y^{\zeta-1}dy > \eta, 1-\gamma < X_{(k)}/b(n/k) < 1+\gamma] + o(1)\\
\le&P[\int_M^\infty \frac{1}{k}\sum_{i=1}^n \epsilon_{X_i/b(n/k)}((1-\gamma)y,\infty]\zeta y^{\zeta-1}dy > \eta] + o(1).
\end{aligned}$$

Ignore the term $o(1)$. Markov's inequality gives a bound

$$\begin{aligned}
\le&(const)\int_M^\infty E\Big(\frac{1}{k}\sum_{i=1}^n P[X_i \ge b(n/k)(1-\gamma)y]\Big)\zeta y^{\zeta-1}dy\\
=&(const)\int_M^\infty \frac{n}{k}\bar F(b(n/k)(1-\gamma)y])\zeta y^{\zeta-1}dy.
\end{aligned}$$

and applying Karamata's theorem [Resnick, 2007, Bingham et al., 1987, Geluk and de Haan, 1987, de Haan, 1970], we have as $n\to\infty$ that this converges to

$$=(const)\int_M^\infty \big((1-\gamma)y\big)^{-\alpha}\zeta y^{\zeta-1}dy \quad \overset{M\to\infty}{\to} \quad 0,$$

as required. This finishes Step 3 and completes the proof modulo the verification that (4.5.8) can be proven for this problem.

The remaining task of checking (4.5.8) proceeds as follows. Recall μ_X^M and μ_Y^M from (4.6.7) and (4.6.9). Fix M. Then for p_δ^n in (4.5.8), we have

$$\begin{aligned} p_\delta^n = \frac{1}{k^M}\#\Bigg\{ j : & \mu_X^M - \delta < -\log\frac{j}{k} < \mu_X^M + \delta, \\ & 0 < -\log\frac{j}{k} \le M; \mu_Y^M - \delta < \log\frac{X_{(j)}}{X_{(k)}} < \mu_Y^M + \delta, \\ & 0 \le \log\frac{X_{(j)}}{X_{(k)}} \le \frac{2M}{\alpha}\Bigg\}. \end{aligned}$$

Since $\mu_X^M \approx 1$ and $\mu_Y^M \approx 1/\alpha$, we get for large M

$$\begin{aligned} p_\delta^n :=& \frac{1}{k^M}\#\Big\{ j : \mu_X^M - \delta < -\log\frac{j}{k} < \mu_X^M + \delta;\ \mu_Y^M - \delta < \log\frac{X_{(j)}}{X_{(k)}} < \mu_Y^M + \delta\Big\} \\ =& \frac{1}{k^M}\#\Big\{ j : \frac{X_{(\lceil k\exp\{-(\mu_X^M - \delta)\}\rceil)}}{X_{(k)}} \vee e^{\mu_Y^M - \delta} < \frac{X_{(j)}}{X_{(k)}} < \frac{X_{(\lceil k\exp\{-(\mu_X^M + \delta)\}\rceil)}}{X_{(k)}} \wedge e^{\mu_Y^M + \delta}\Big\} \\ =& \frac{k}{k^M}\,\hat{\nu}_n\Bigg(\frac{X_{(\lceil k\exp\{-(\mu_X^M - \delta)\}\rceil)}}{X_{(k)}} \vee e^{\mu_Y^M - \delta}, \frac{X_{(\lceil k\exp\{-(\mu_X^M + \delta)\}\rceil)}}{X_{(k)}} \wedge e^{\mu_Y^M + \delta}\Bigg). \end{aligned}$$

Apply (4.6.4) and (4.6.5) and we find

$$p_\delta^n \xrightarrow{P} \frac{1}{1 - e^{-M}}\nu_\alpha\Big(e^{(\mu_X^M - \delta)/\alpha} \vee e^{\mu_Y^M - \delta}, e^{(\mu_X^M + \delta)/\alpha} \wedge e^{\mu_Y^M + \delta}\Big).$$

Since $\mu_X^M \approx 1$ and $\mu_Y^M \approx 1/\alpha$, by picking M large and δ small, the right side can be made to be less than 1. This completes the proof. $\square$

APPENDIX A

APPENDIX

A.1 Appendix

For convenience, this section collects some notation, needed background on regular variation and notions on vague convergence needed for some formulations and proofs.

A.1.1 Regular variation and the function classes Π

Regular variation is the mathematical underpinning of heavy tail analysis and extreme value theory. This topic has been discussed nicely in Resnick [2007, 2008b], Seneta [1976], Geluk and de Haan [1987], de Haan [1970], de Haan and Ferreira [2006], Bingham et al. [1987] among others.

A measurable function $U(\cdot) : \mathbb{R}_+ \to \mathbb{R}_+$ is regularly varying at ∞ with index $\rho \in \mathbb{R}$, denoted by $U \in RV_\rho$, if for $x > 0$,

$$\lim_{t\to\infty} \frac{U(tx)}{U(t)} = x^\rho. \tag{A.1.1}$$

Regular variation is connected closely to the concept of extreme value domain of attraction of a distribution function. Let a distribution function $F \in D(G_\gamma)$, G_γ as defined in (1.1.1). In general this means there exist functions $a(t) > 0, b(t) \in \mathbb{R}$, such that,

$$F^t(a(t)y + b(t)) \to G_\gamma(y), \quad (t \to \infty), \tag{A.1.2}$$

weakly, where

$$G_\gamma(y) = \exp\{-(1+\gamma y)^{-1/\gamma}\}, \quad 1+\gamma y > 0,\ \gamma \in \mathbb{R}, \tag{A.1.3}$$

and the expression on the right is interpreted as $e^{-e^{-y}}$ if $\gamma = 0$. See, for example, Resnick [2008b], Embrechts et al. [1997], de Haan [1970], Coles [2001], Reiss and Thomas [2001]. In terms of regular variation this means that for $\gamma > 0$, we have the right tail of F to be regularly varying,that is, $\bar{F} = 1 - F \in RV_{-1/\gamma}$.

We can and do assume

$$b(t) = \Big(\frac{1}{1-F(\cdot)}\Big)^{\leftarrow}(t).$$

Thus, we have relation (A.1.2) is equivalent to

$$t\bar{F}(a(t)y + b(t)) \to (1+\gamma y)^{-1/\gamma}, \quad 1+\gamma y > 0, \tag{A.1.4}$$

or taking inverses, as $t \to \infty$,

$$\frac{b(ty) - b(t)}{a(t)} \to \begin{cases} \frac{y^\gamma - 1}{\gamma}, & \text{if } \gamma \neq 0, \\ \log y, & \text{if } \gamma = 0. \end{cases} \tag{A.1.5}$$

In such a case we say that $b(\cdot)$ is *extended regularly varying* with auxiliary function $a(\cdot)$ and we denote $b \in ERV_\gamma$. In case $\gamma = 0$, we say that $b(\cdot) \in \Pi(a(\cdot))$; that is, the function $b(\cdot)$ is Π-*varying* with auxiliary function $a(\cdot)$ [Resnick, 2008b, Bingham et al., 1987, de Haan and Ferreira, 2006].

More generally (de Haan and Resnick [1979], de Haan and Ferreira [2006]) define for an auxiliary function $a(t) > 0$, $\Pi_+(a)$ to be the set of all functions $\pi : \mathbb{R}_+ \mapsto \mathbb{R}_+$ such that

$$\lim_{t\to\infty} \frac{\pi(tx) - \pi(t)}{a(t)} = k \log x, \quad x > 0,\ k > 0. \tag{A.1.6}$$

The class $\Pi_-(a)$ is defined similarly except that $k < 0$ and

$$\Pi(a) = \Pi_+(a) \cup \Pi_-(a).$$

By adjusting the auxiliary function in the denominator, it is always possible to assume $k = \pm 1$.

Two functions $\pi_i \in \Pi_\pm(a)$, $i = 1, 2$ are $\Pi(a)$-equivalent if for some $c \in \mathbb{R}$

$$\lim_{t\to\infty} \frac{\pi_1(t) - \pi_2(t)}{a(t)} = c.$$

There is usually no loss of generality in assuming $c = 0$. The following are known facts about Π-varying functions.

1. We have $\pi \in \Pi_+(a)$ iff $1/\pi \in \Pi_-(a/\pi^2)$.
2. If $\pi \in \Pi_+(a)$, then ([de Haan and Resnick, 1979, page 1031] or [Bingham et al., 1987, page 159]) there exists a continuous and strictly increasing $\Pi(a)$-equivalent function π_0 with $\pi - \pi_0 = o(a)$.
3. If $\pi \in \Pi_+(a)$, then
$$\lim_{t\to\infty} \pi(t) =: \pi(\infty)$$
exists. If $\pi(\infty) = \infty$, then $\pi \in RV_0$ and $\pi(t)/a(t) \to \infty$. If $\pi(\infty) < \infty$, then $\pi(\infty) - \pi(t) \in \Pi_-(a)$ and $\pi(\infty) - \pi(t) \in RV_0$ and $(\pi(\infty) - \pi(t))/a(t) \to \infty$. (Cf. [Geluk and de Haan, 1987, page 25].) Furthermore,
$$\frac{1}{\pi(\infty) - \pi(t)} \in \Pi_+\big(a/(\pi(\infty) - \pi(t))^2\big).$$

A.1.2 Tail equivalence

Suppose $\boldsymbol{X}$ and $\boldsymbol{Y}$ are $\mathbb{R}_+^d$-valued random vectors. Then $\boldsymbol{X}$ and $\boldsymbol{Y}$ are *tail equivalent* (in the context of multivariate regular variation) on a cone $\mathfrak{C} \subset \overline{R}_+^d$ if there exists a scaling function $b(t) \uparrow \infty$ such that

$$t\mathbf{P}\Big(\frac{\boldsymbol{X}}{b(t)} \in \cdot\Big) \xrightarrow{v} \nu(\cdot) \qquad \text{and} \qquad t\mathbf{P}\Big(\frac{\boldsymbol{Y}}{b(t)} \in \cdot\Big) \xrightarrow{v} c\nu(\cdot)$$

in $M_+(\mathfrak{C})$ for some $c > 0$ and non-null Radon measure ν on $\mathfrak{C}$. We denote this concept by $X \overset{te(\mathfrak{C})}{\sim} Y$. See Maulik and Resnick [2005].

A.1.3 Vague convergence

For a nice space $\mathbb{E}^*$, that is, a space which is locally compact with countable base (for example, a finite dimensional Euclidean space), denote $\mathbb{M}_+(\mathbb{E}^*)$ for the non-negative Radon measures on Borel subsets of $\mathbb{E}^*$. This space is metrized by the vague metric. The notion of vague convergence in this space is as follows: If $\mu_n \in \mathbb{M}_+(\mathbb{E}^*)$ for $n \geq 0$, then μ_n converge vaguely to μ_0 (written $\mu_n \xrightarrow{v} \mu_0$) if for all bounded continuous functions f with compact support we have

$$\int_{\mathbb{E}^*} f d\mu_n \to \int_{\mathbb{E}^*} f d\mu_0 \quad (n \to \infty).$$

This concept allows us to write (2.1.2) as

$$tP[\frac{Y - b(t)}{a(t)} \in \cdot\,] \xrightarrow{v} m_\gamma(\cdot), \tag{A.1.7}$$

vaguely in $\mathbb{M}_+((-\infty, \infty])$ where

$$m_\gamma((x, \infty]) = (1 + \gamma x)^{-1/\gamma}.$$

Standard references include Kallenberg [1983], Neveu [1977] and [Resnick, 2008b, Chapter 3].

BIBLIOGRAPHY

G. Beer. *Topologies on Closed and Closed Convex Sets*, volume 268 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1993. ISBN 0-7923-2531-1.

J. Beirlant, P. Vynckier, and J. Teugels. Tail index estimation, Pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc.*, 91(436):1659–1667, 1996. ISSN 0162-1459.

P. Billingsley. *Convergence of Probability Measures*. John Wiley & Sons Inc., New York, 1968.

P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. ISBN 0-471-00710-2. A Wiley-Interscience Publication.

N.H. Bingham, C.M. Goldie, and J.L. Teugels. *Regular Variation*. Cambridge University Press, 1987.

S.G. Coles. *An Introduction to Statistical Modeling of Extreme Values.* Springer Series in Statistics. London: Springer. xiv, 210 p. , 2001.

S.G. Coles and J.A. Tawn. Modelling extreme multivariate events. *J. R. Statist. Soc. B*, 53:377–392, 1991.

S.G. Coles, J.E. Heffernan, and J.A. Tawn. Dependence measures for extreme value analyses. *Extremes*, 2(4):339–365, 1999.

S. Csörgő, P. Deheuvels, and D. Mason. Kernel estimates for the tail index of a distribution. *Ann. Statist.*, 13:1050–1077, 1985.

Y. Davydov, I. Molchanov, and S. Zuyev. Stable distributions and harmonic analysis on convex cones. *C. R. Math. Acad. Sci. Paris*, 344(5):321–326, 2007. ISSN 1631-073X.

L. de Haan. *On Regular Variation and Its Application to the Weak Convergence of Sample Extremes*. Mathematisch Centrum Amsterdam, 1970.

L. de Haan. A characterization of multidimensional extreme-value distributions. *Sankhyā Ser. A*, 40(1):85–88, 1978. ISSN 0581-572X.

L. de Haan and J. de Ronde. Sea and wind: multivariate extremes at work. *Extremes*, 1(1):7–46, 1998.

L. de Haan and A. Ferreira. *Extreme Value Theory: An Introduction*. Springer-Verlag, New York, 2006.

L. de Haan and S.I. Resnick. Limit theory for multivariate sample extremes. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 40:317–337, 1977.

L. de Haan and S.I. Resnick. Conjugate π-variation and process inversion. *Ann. Probab.*, 7(6):1028–1035, 1979. ISSN 0091-1798.

P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extreme Events for Insurance and Finance*. Springer-Verlag, Berlin, 1997.

P. Feigin and S.I. Resnick. Linear programming estimators and bootstrapping for heavy tailed phenomena. *Adv. in Appl. Probab.*, 29:759–805, 1997.

J. Flachsmeyer. Verschiedene topologisierungen im Raum der abgeschlossenen Mengen. *Math. Nachr.*, 26:321–337, 1963/1964. ISSN 0025-584X.

A. Fougères and P. Soulier. Modelling and estimation of conditional excesses. 2008. URL http://arXiv.org/0806.2426.

J.L. Geluk and L. de Haan. *Regular Variation, Extensions and Tauberian Theorems*, volume 40 of *CWI Tract*. Stichting Mathematisch Centrum, Centrum voor Wiskunde en Informatica, Amsterdam, 1987. ISBN 90-6196-324-9.

J.E. Heffernan and S.I. Resnick. Hidden regular variation and the rank transform. *Adv. Appl. Prob.*, 37(2):393–414, 2005.

J.E. Heffernan and S.I. Resnick. Limit laws for random vectors with an extreme component. *Ann. Appl. Probab.*, 17(2):537–571, 2007. ISSN 1050-5164. doi: 10.1214/105051606000000835.

J.E. Heffernan and J.A. Tawn. A conditional approach for multivariate extreme values (with discussion). *JRSS B*, 66(3):497–546, 2004.

B.M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, 3:1163–1174, 1975.

O. Kallenberg. *Random Measures*. Akademie-Verlag, Berlin, third edition, 1983. ISBN 0-12-394960-2.

C. Klüppelberg and S.I. Resnick. The Pareto copula, aggregation of risks and the emperor's socks. *Journal of Applied Probability*, 45(1):67–84, 2008.

M. Kratz and S.I. Resnick. The qq–estimator and heavy tails. *Stochastic Models*, 12:699–724, 1996.

A.W. Ledford and J.A. Tawn. Statistics for near independence in multivariate extreme values. *Biometrika*, 83(1):169–187, 1996. ISSN 0006-3444.

A.W. Ledford and J.A. Tawn. Modelling dependence within joint tail regions. *J. Roy. Statist. Soc. Ser. B*, 59(2):475–499, 1997. ISSN 0035-9246.

A.W. Ledford and J.A. Tawn. Concomitant tail behaviour for extremes. *Adv. in Appl. Probab.*, 30(1):197–215, 1998. ISSN 0001-8678.

D. Mason. Laws of large numbers for sums of extreme values. *Ann. Probab.*, 10: 754–764, 1982.

D. Mason and T. Turova. Weak convergence of the Hill estimator process. In J. Galambos, J. Lechner, and E. Simiu, editors, *Extreme Value Theory and Applications*, pages 419–432. Kluwer Academic Publishers, Dordrecht, Holland, 1994.

G. Matheron. *Random Sets and Integral Geometry*. John Wiley & Sons, New York-London-Sydney, 1975. With a foreword by G.S. Watson, Wiley Series in Probability and Mathematical Statistics.

K. Maulik and S.I. Resnick. Characterizations and examples of hidden regular variation. *Extremes*, 7(1):31–67, 2005.

K. Maulik, S.I. Resnick, and H. Rootzén. Asymptotic independence and a network traffic model. *J. Appl. Probab.*, 39(4):671–699, 2002. ISSN 0021-9002.

A.J. McNeil, R. Frey, and P. Embrechts. *Quantitative Risk Management*. Princeton Series in Finance. Princeton University Press, Princeton, NJ, 2005. ISBN 0-691-12255-5. Concepts, techniques and tools.

T. Mikosch. How to model multivariate extremes if one must? *Statist. Neerlandica*, 59(3):324–338, 2005. ISSN 0039-0402.

T. Mikosch. Copulas: Tales and facts. *Extremes*, 9(1):3–20, 2006. ISSN 1386-1999 (Print) 1572-915X (Online).

I. Molchanov. *Theory of Random Sets*. Probability and its Applications (New York). Springer-Verlag London Ltd., London, 2005. ISBN 978-185223-892-3; 1-85233-892-X.

J. Neveu. Processus ponctuels. In *École d'Été de Probabilités de Saint-Flour, VI—1976*, pages 249–445. Lecture Notes in Math., Vol. 598, Berlin, 1977. Springer-Verlag.

J. Pickands. Statistical inference using extreme order statistics. *Ann. Statist.*, 3: 119–131, 1975.

J. Pickands. Multivariate extreme value distributions. In *43rd Sess. Int. Statist. Inst.*, pages 859–878, 1981.

R.D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values*. Birkhäuser Verlag, Basel, second edition, 2001. ISBN 3-7643-6487-4.

S.I. Resnick. *A Probability Path*. Birkhäuser, Boston, 1999.

S.I. Resnick. Hidden regular variation, second order regular variation and asymptotic independence. *Extremes*, 5(4):303–336, 2002.

S.I. Resnick. Modeling data networks. In B. Finkenstadt and H. Rootzén, editors, *SemStat: Seminaire Europeen de Statistique, Extreme Values in Finance, Telecommunications, and the Environment*, pages 287–372. Chapman-Hall, London, 2003.

S.I. Resnick. Multivariate regular variation on cones: application to extreme values, hidden regular variation and conditioned limit laws. *Stochastics: An International Journal of Probability and Stochastic Processes*, 80:269–298, 2008a. ISSN 1744-2508.

S.I. Resnick. *Heavy Tail Phenomena: Probabilistic and Statistical Modeling*. Springer Series in Operations Research and Financial Engineering. Springer-Verlag, New York, 2007. ISBN: 0-387-24272-4.

S.I. Resnick. *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008b. ISBN 978-0-387-75952-4. Reprint of the 1987 original.

S.I. Resnick and C. Stărică. Consistency of Hill's estimator for dependent data. *J. Appl. Probab.*, 32(1):139–167, 1995. ISSN 0021-9002.

S. Sarvotham, R. Riedi, and R. Baraniuk. Network and user driven on-off source model for network traffic. *Computer Networks*, 48:335–350, 2005. Special Issue on "Long-range Dependent Traffic".

E. Seneta. *Regularly Varying Functions*. Springer-Verlag, New York, 1976. Lecture Notes in Mathematics, 508.

R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons Inc., New York, 1980. ISBN 0-471-02403-1.

N.V. Smirnov. Limit distributions for the terms of a variational series. *Trudy Mat. Inst. Steklov.*, 25:60, 1949. ISSN 0371-9685.

W. Vervaat. Random upper semicontinuous functions and extremal processes. In W. Vervaat and H. Holwerda, editors, *Probability and Lattices*, volume 110 of *CWI Tract*, pages 1–56. Math. Centrum Wisk. Inform., Amsterdam, 1997.